Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Responses to Material Presented During Various Levels of Sleep: C. W. SIMON AND W. H. EMMONS 89

The Precision of Choice Reactions: E. C. POULTON 98

Effect of Long-Term Practice and Time-on-Target Information Feedback on a Complex Tracking Task: E. J. ARCHER, G. W. KENT, AND F. A. MOTE 103

Facilitating Effects of Induced Tension Upon the Perception Span for Digits: W. A. SHAW 113

Acquisition of Motor Skill: IV. Effects of Repeated Periods of Massed Practice: R. B. AMMONS AND L. WILLIS 118

Level of Conditioning and Intensity of the Adaptation Stimulus: J. A. TAYLOR 127

The Correlates of Manifest Anxiety in Stylus Maze Learning: H. S. AXELROD, E. L. COWEN, AND F. HEILIZER 131

Development and Maintenance of the Preference Value of an Object: H. M. SCHROEDER ... 139

The Effect of Ego Orientation and Problem Difficulty on Muscle Action Potentials: M. E. REUBER 142

Retroactive Inhibition as a Function of the Temporal Position of the Interpolated Learning: J. M. NEWTON AND D. D. WICKENS 149

Spontaneous Recovery of R_A Following Interpolated Acquisition and Extinction of R_B: A. E. HICKEY, JR. 155

Individual Behavior in Social Situations: Its Relation to Anxiety, Neuroticism, and Group Solidarity: V. CERVIN 161

American Psychological Association

Vol. 51 No. 2 February 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S's*, and *Ss'*; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright © 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 51, NO. 2 FEBRUARY, 1956

RESPONSES TO MATERIAL PRESENTED DURING VARIOUS LEVELS OF SLEEP

CHARLES W. SIMON AND WILLIAM H. EMMONS[1]

The RAND Corporation

Recently, there has been an increased interest in the possibility of learning during sleep. A critical review of the few scientific studies in this field to date lead to the conclusion that the evidence supporting claims of learning during actual sleep is inconclusive. The chief criticism against the existing studies is their failure to continuously determine the sleep state of *Ss* during the stimulus input period (**16**).

After considerable exploratory work (**15**), the present experiment was designed to study the effect of presenting material at different levels between wakefulness and deep sleep on the ability of *Ss* (*a*) to respond to it immediately, and (*b*) to recall it later upon awakening. It was hypothesized that learning during sleep was improbable.

METHOD

Subjects

Twenty-one male experimental *Ss* were used. In order to facilitate the detection of gradual changes which occur in the EEG between wakefulness and light sleep, it was necessary to use only *Ss* having a persistent occipital alpha rhythm when awake and relaxed with their eyes closed. Ten *Ss* were junior college students, nine were scientists, and two were policemen. Results from students and policemen were combined since they were similar on critical variables. Means and *SD*'s of descriptive variables for the experimental subgroups were determined. For the scientist group, these values were: age, 30.6 yr. ±5.0; IQ, 122.6 ± 6.0; items correct on pretest, 11.7 ± 3.4. For the college men and policemen combined, these values were: age, 23.2 yr. ± 6.3; IQ, 108.6 ± 4.7; items correct on pretest, 6.8 ± 3.0.

Forty junior college males and 24 male scientists acted as a control group without regard for their alpha rhythm. These *Ss* were used to obtain an estimate of the probability of correct answers being chosen on a multiple-choice test, since it seemed unlikely that educated *Ss* would select their answers to unknown questions solely by chance. This control group was first given a pretest to determine the number of test questions known without training and before seeing the multiple-choice alternatives. With no intervening training and after items correct on the pretest were removed, the pretest significantly predicted scores on the multiple-choice test. This prediction was not significantly improved when IQ or age was added in a multiple-regression equation.

Control and experimental subgroups were matched on means and variances of their IQ, age, and pretest scores. Only the variances of the age and IQ of the college subgroups were unmatched.[2]

[1] The authors wish to acknowledge the valuable assistance of Louie W. Mason, Jr. and James L. Barnes in the analysis of the data in this study. Appreciation is expressed to the faculty and students of Santa Monica City College, to the Santa Monica Police Department, and to the volunteer *Ss* within The RAND Corporation for their cooperation throughout the project.

[2] Supporting and additional statistics along with data by *S* have been deposited with the

Apparatus

The *Ss* slept in clean comfortable beds in three separate soundproof, air conditioned, electrically-shielded booths. EEG electrodes were applied to the right occipital area, the vertex, and the left mastoid process. The electrode wires were arranged to allow relatively free movement and comfort during sleep. Two monopolar EEG recordings (right occipital and vertex) were made from each *S* using a six-channel Offner electroencephalograph and Dynograph inkwriter. A marker pen mounted on the inkwriter automatically marked the exact time an auditory stimulus was presented to *S*. The learning material was recorded on magnetic tape and played through loud-speakers placed inside the booths. A two-way intercommunication system allowed *Es* to communicate with *Ss*.

Procedure

Pretesting period.—Several hours before retiring, *Ss* were pretested to discover what information they knew in order to have a base from which to evaluate how much had been learned following "sleep-training" on the same material. They were encouraged to guess answers to as many questions as they could, since a chance guess might be the correct one. No time limit was imposed.

Ninety-six general information questions were culled from various books on the basis of the following criteria: (*a*) the information was not generally known; (*b*) the answers required were not outside the verbal experience of the *Ss*; (*c*) questions could be answered in short phrases or single words. These same 96 questions and their answers were recorded on tape at 5-min. intervals for presentation during the sleep period. The answer was presented as a restatement of the question with the critical information near the end. For example, Question: "In what kind of store did Ulysses S. Grant work before the war?" Answer: "Before the war, Ulysses S. Grant worked in a hardware store."

Preliminary period.—Following the pretest, *Ss* prepared for bed and the EEG electrodes were applied. Three *Ss* were run simultaneously. Since it was not possible to adjust the auditory intensity of the stimulus material in each booth individually, it was necessary to find a level satisfactory to all three. "Satisfactory" was defined as the point at which the training material could be heard clearly by the waking *S*, yet not be loud enough to materially disturb him once he was asleep. As *Ss* lay in their booths, practice verbal material was played at approximately the same loudness as the experimental stimulus material was to be played. This loudness level was adjusted after a number of tests so that the *S* with the highest threshold was able to hear the words clearly.

Before turning out the lights and allowing *Ss* to go to sleep, the following instructions were given: (*a*) if they awakened during the night at any time, they were to call out their name and booth number as soon as possible and to say that they were awake; (*b*) if they heard the answer to any of the question-answer combinations during the night, they were to wait until the answer was completed, and then call out their name and booth number. In order to impress these instructions on *S*, each was asked to repeat them back to *E*.

Training period.—Within 5-min. after *S* retired, the tape recorder was turned on and the question-answer combinations were played into the booths at approximately 5-min. intervals. A pen automatically marked the EEG record whenever a question or answer was played. The *Es* marked on the record whether or not *S* stated he heard the answer. Records of any other pertinent information were kept, e.g., if *S* awoke, his remarks, and so forth.

Postiest period.—Shortly after the last question and answer was played, *Ss* were awakened (if they were not already so), permitted to dress, and wash their face and hands. The same questions which they received on the pretest and during the training were again presented for them to answer; as before, *Ss* were encouraged to guess. This unaided recall was followed by a multiple-choice test in which five alternative answers (including the correct one) were provided with the questions.

Control group testing.—All members of the control groups were given the 96 questions and asked to answer them by guessing whenever they were not certain. Following a brief rest and no training, they were given the questions along with the multiple-choice answers and retested.

Intelligence testing.—Following the completion of the experiment, both experimental and control groups were given the 30-min. Otis Self-Administering Test of Mental Ability, Form D.

Results

The basic data available for analysis were: (*a*) the levels of sleep during the training period as indicated by EEG records, (*b*) the frequencies with which *Ss* immediately reported they had heard an answer, and (*c*) the frequencies with which *Ss* correctly recalled an answer, both unaided and by

American Documentation Institute. Order Document No. 4738, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 × 8 in. photocopies.

TABLE 1

EEG AND PSYCHOLOGICAL CONDITIONS ALONG THE SLEEP–WAKEFULNESS CONTINUUM

Level	% Items	% Ss	Electroencephalogram	Psychological Condition
O	8	90	Continuous or nearly continuous alpha of maximum amplitude and frequency of not more than 1 cps slower than the *S*'s normal alpha frequency. Frequency and amplitude are slightly less after sleeping than before.	Awake. Relaxed with eyes closed. Responsive to external stimulation.
A+	9	100	More than 50% of the scoring period contains alpha. Also low level, random activity characteristic of an alpha block may be present with alpha disappearing at the onset of stimulation and returning shortly after its cessation.	Drowsy.
A	12	95	Less than 50% alpha but scoring period contains at least three cycles of activity having the same frequency as the O level. The alpha amplitude may be of the same or considerably lower amplitude than before.	Attention wanders; reverie. Increased reaction time.
A−	14	100	Contains cyclical activity having a frequency more than 1 cps slower than the Level O record. May include waves of mixed duration with periods between .12 and .08 sec. with no one period being dominant. Also includes records showing no alpha rhythm during stimulation, but with alpha occurring within 30 sec. prior to stimulation or following but not both.	Partial awareness.
B	22	100	Absence of alpha during the stimulus period and the adjacent 30 sec. of record. Stimulus effects may occur consisting of low level fast activity or an increase in activity containing both high and low frequency components, with some waves having periods within the alpha range. Low-level delta activity is present in the absence of stimulus effects.	Transition. Dreamlike state. Infrequent responses to external stimulation. Onset of sleep. Easily awakened.
C	18	100	Absence of alpha with an increase of delta and the appearance of sleep spindles (14 cps). This state is characterized by stimulus effects such as increases in amplitude of delta waves with the onset of the stimulus. Type of effects vary with individuals.	Light sleep. No behavioral responses to external stimulation (unless stimulus awakens). Dreams sometimes remembered.
D	7	90	A further increase in delta amplitude with a reduction in frequency and diminution of stimulus effects and sleep spindles. Amplitude of delta almost at maximum.	Deep sleep. No memory for dreams. Difficult to awaken.
E	3	57	Absence of sleep spindles and stimulus effects. Very large delta activity with smooth waves of .5- to 1.5-sec. duration.	Very deep sleep.

recognizing the correct multiple-choice alternative.

Sleep levels.—Those portions of the EEG record occurring simultaneously with the presentation of the answers were categorized into 22 distinguishable patterns. These 22 EEG patterns were divided into eight groups, or sleep levels, which were ordered on their proportion of visually observable cyclical activity within the alpha frequency range (8–13 cps) and on their

amplitude, frequency, and the effects of stimulation when alpha frequencies were no longer present. This order corresponds to *S*'s state of consciousness between wakefulness and deep sleep and is described in Table 1.

The categories were based on *Ss* having continuous or nearly continuous waking alpha and read primarily from the monopolar occipital record. The letter symbols and the descriptions of the sleep levels in Table 1 approximate those of other experimenters in the field. Most investigators believe that an individual is asleep when Level C is reached, although some believe the onset of sleep occurs toward the end of Level B (3, 5, 9, 11, 12).

The EEG records were analyzed independently by two *Es*, after first agreeing on a few major classification principles. Their independent readings of the eight levels correlated .877 (corrected contingency coefficient) with $\sigma_r = \pm .024$ when the true correlation is zero. Following this first reading, the two *Es* worked together to agree on the categories for all 2,016 items, elaborating on the principles for such categorization and compromising in a few cases where it was necessary. The breakdown of EEG levels during the drowsy stages before sleep is finer in this paper than has been reported previously. In the final analysis of the 2,016 sections of record, 7% were unclassifiable due to artifacts which obscured the EEG pattern.

Levels O through A will be designated as "alpha levels" and B through E as "non-alpha levels." Cyclical activity within the alpha frequency range (8–13 cps) is found in the first four levels and is absent in the second four, with the exception of highly questionable alpha frequencies appearing during stimulation at the borderline between waking and sleeping. Sleep levels always refer to the levels at the time of stimulus presentation and never at the time of the immediate response or of later recall. All scoring of the EEG records was completed prior to the scoring of the information tests. When there was doubt as to the presence of alpha due to its low amplitude, very small quantity, or masking by other activity, it was the policy to place these patterns into the deepest alternative level.

Immediate response and unaided recall.—The "immediate response" or "reported heard" analysis was based on 1,827 items; items were removed when the EEG was obscured by artifacts and when it was impossible to determine which of the three *Ss* running simultaneously had responded. The "unaided recall" analysis was based on 1,690 items; items were removed when the EEG was obscured by artifacts and when the items were known on the pretraining test.

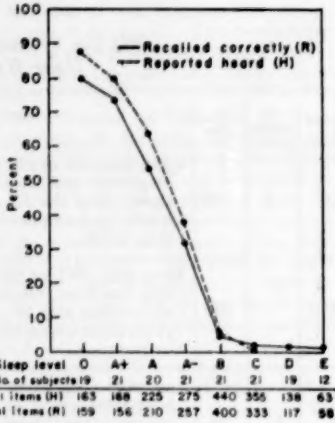

FIG. 1. The percentage of items reported heard during the stimulation at varying levels along the wake-sleep continuum and the percentage subsequently recalled correctly unaided. The number of *Ss* contributing to each level and the number of items from which each percentage was computed are shown.

The percentage of items reported heard at time of presentation and the percentage correctly recalled unaided, upon awakening, are shown in Fig. 1. These percentages in Fig. 1, based on the combined data for all *Ss* and items occurring at each level, decrease as the amount of alpha decreases and delta appears, that is, as the sleep level deepens.

Of the items recalled correctly on the entire test, the following percentages were reported heard: Level O, 90%; Level A+, 85%; Level A, 79%; Level A−, 69%; Level B, 49%. Below Level B no items were reported heard and only 8 out of 508 items were answered correctly. Four *Ss* indicated that after an extended period of wakefulness, they heard answers but

TABLE 2

NUMBER OF ITEMS REPORTED HEARD (H−), ANSWERED CORRECTLY (R−), OR BOTH (HR) IN THE NON-ALPHA LEVELS

Level	Items Possible	HR	H−	R−	No. of *Ss* Contributing
B	400	8	18*	11**	15
C	333	0	0	5	4
D	117	0	0	2	2
E	58	0	0	1	1
All	908	8	18	19	

* Five of these were answered correctly on the pretest.

** Two of these may or may not have been heard; answers were too indistinct to determine which of three *Ss* reported.

deliberately failed to report since they wanted to go to sleep.

The percentage correct at each level when based on the combined data did not differ significantly from the percentage correct computed as the mean of all *Ss*' individual percentage correct ($t = 1.82$, 7 *df*). All individual *Ss* showed the same tendency to reduce the percentage heard and recalled as alpha disappeared.

Some items, however, were immediately responded to and recalled later in Levels B through E. The breakdown of these items by level, type of response, and number of *Ss* contributing are shown in Table 2. Twelve of the 19 correctly recalled items presented in Level B appear to be the effect of bona fide learning. These were cases when the Level B EEG patterns were of specific types occurring during stimulation and were related to arousal from sleep. These patterns were characterized by low level, fast random activity quite similar to the "activation pattern" found in animals (**10, 14**), as well as by an increase in random activity containing both high and low frequency components. In the latter pattern, mixed alpha frequencies occurred simultaneously with significant quantities of delta activity. Both patterns appeared to be different aspects of a stimulus effect at the borderline between sleep and wakefulness. Since both alpha and delta appeared, the pattern was assigned to Level B, the transition state, rather than Level A− in accordance with the policy stated earlier. When these patterns are observed, arousal is imminent if stimulation continues. All of the items reported heard and recalled correctly (HR), half of the items reported heard only (H−), and four of the items recalled correctly (R−) in Level B could be accounted for by these patterns associated with arousal.

Eight items in Table 2 were heard only without any arousal patterns being observed. It cannot be determined from this study whether the answers which were reported heard were also understood; at least, they were not recalled.

Were the remaining items recalled correctly in Levels B through E (after removing items occurring during the arousal patterns) really learned during sleep-training? Since the range of reasonable answers to many of the questions in this experiment was restricted, the likelihood of a correct guess on the posttest increased if an *S* habitually changed his answer from first to second trial. Although it can never be proven that these 15 out of 508 items were answered correctly by guessing, the hypothesis was tested as follows: An index of changeability was determined for each *S* by examining only those items which he answered incorrectly on both pre- and posttests; this was done in order to avoid changes brought about by obvious learning. When the number of items recalled correctly without being reported heard and not associated with arousal EEG patterns was correlated with the percentage of incorrect answers changed between pre- and posttests by the 21 *Ss*, the relationship was positive

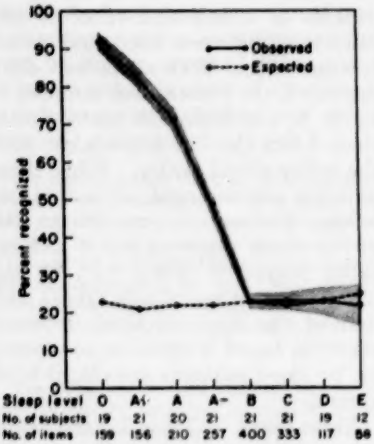

FIG. 2. The percentage of answers recognized on the multiple-choice test after being presented at varying levels along the continuum between a waking and deep sleep state. The shaded portion represents plus and minus one standard error of the percentage. The expected value was that obtained from an untrained control group of comparable ability answering the same items.

($r = .68$). This correlation, significant below the .01 probability level, supports the hypothesis that those persons changing their answers the most increased the number of correct guesses that they made. In other words, the 15 correct items could be the result of guessing.

Examination of individual *S*'s data revealed that they all followed the same trend as shown by the curves in Fig. 1. The few cases of apparent recall that are shown in Table 2, accounting for roughly 1.5% of all the items below Level B, are attributable to only a few *S*s. No *S* contributed more than two items and their occurrence in any sleep level was haphazard.

Recognition test.—Figure 2 shows the percentage of items recalled correctly for all *S*s when multiple-choice answers were provided, along with the standard error of these percentages. As in Fig. 1, with a decrease in alpha frequencies at the time of presentation, there was a corresponding decrease in the number of items recalled later. Unlike the others, however, the recall curve from this recognition test leveled off around 23% instead of dropping to zero.

Since it was possible for *S*s to answer correctly solely by chance on the multiple-choice test, it was necessary to compare the percentages obtained by the experimental groups with an empirical estimate of what the theoretical percentage would be. Although 20% was chance expectancy for this test, the difficulty of making a perfect multiple-choice test with answers equal in selectivity caused the *E*s to use scores of a control group as the basis for estimating the expected value.

It was found that the scientist control subgroup answered 26% correctly without training and the college control subgroup answered 22% correctly. The expected percentages shown in Fig. 2 were based on empirical values from the control groups weighted according to the items contributed by the scientist and college-police experimental groups in each sleep level.

Thus, from Fig. 2, it can be seen that the expected values remain essentially the same for all sleep levels, while the values observed from the experimental group changed. In Levels O through A−, the experimental *S*s showed a considerably higher percentage correct, while in Levels B through E, the percentage correct for the experimental group was essentially equal to (within $\pm.5\sigma_{\%}$) and in most cases less than the mean expected value.

Since the values in Fig. 2 were based on combined data for all *S*s, hiding possible individual learning, the data

were analyzed by *S*. In order to have a sufficiently large expected value for computing χ^2 for each *S*, level O through A− (alpha levels) were combined as were Levels B through E (non-alpha levels). These were compared with the mean percentage correct for the appropriate control subgroup. The χ^2s from these analyses are shown in Table 3. No experimental *S* recalled correctly significantly more multiple-choice items than would be expected from the estimated percentages obtained from the untrained control group.

In order to determine whether any learning had occurred in the lightest sleep level but had been masked by combining those data with the deeper sleep levels, individual χ^2's were computed whenever possible for the results in Level B. Twelve of the *S*s had scores equal to or less than the expected value. Three *S*s scored insignificantly higher than the expected value (χ^2s = .34, .21, and .01; 1 *df*). For five cases, expected values were so small that χ^2 could not be computed. Two of these, with a 22% expected value, had 5 out of 14 and 4 out of 16 answered correctly while three *S*s, with a 26% expected value, had 4 out of 7, 2 out of 4, and 3 out of 10 answered correctly. Only one *S* showed a significant χ^2 of 5.9, with 1 *df*. for Level B. For this *S*, 11 out of 24 items were answered correctly. Of these 24, eight occurred when EEG patterns associated with arousal occurred; when these items were removed, 5 out of 16 were correct where 4 correct were expected.

TABLE 3

χ^2's of Individual *S*s Showing the Significance * of Differences between Observed and Expected Values of Learnings in Alpha and Non-Alpha Levels

S	Levels O through A−	Levels B through E	*S*	Levels O through A−	Levels B through E
1	10.38	.14	12	110.89	.55
2	33.12	.40	13	37.27	.67
3	46.39	3.97**	14	37.51	.08
4	151.30	.39	15	80.58	2.65
5	7.31	.54	16	38.68	.20
6	32.43	2.00	17	32.00	3.57
7	55.69	.15	18	109.24	.84
8	5.41	.05	19	113.78	1.08
9	47.35	.01	20	78.57	.31
10	43.10	1.65	21	6.24	1.9
11	.47	.03			

* χ^2 values higher than 3.84 (1 *df*) indicate that *S* answered significantly more items correct on the multiple-choice test than would be expected from the mean of his control subgroup ($P = .05$).

** This *S* answered 5 out of 49 correctly. This was significantly *less* than the 10.8 expected correct when based on scores of his appropriate control subgroup.

Discussion

The results support the hypothesis that learning during sleep is unlikely. Although a few items were answered correctly unaided, with the easier recognition test (**13**) no more items were answered in Levels B through E by the experimental group than would be expected from the performance of the untrained control group. Since it is generally conceded that sleep occurs somewhere during Level B, then it appears that learning was slight, if any, at this point or below. There appears some basis for assuming that many of these atypical cases were artifacts and not true sleep-learning.

What had been done in the present sleep-learning study that had not been done adequately in any previous studies (**16**) was the careful monitoring of *S*s during the presentation period so that their levels of sleep during training were always known. The use of the EEG for this purpose proved most adequate.

The experiment was designed to get a continuous function rather than a dichotomous relation between responsiveness and consciousness along the wake-sleep continuum. Preliminary exploration (**15**) suggested the present design to be an appropriate one, although it involved certain calculated limitations related to the selection of *S*s, the conditions under which experimental and control groups

were tested, and the number of stimulus repetitions.

Research on alpha dominant *Ss* cannot justifiably be generalized to the general population, although there is no conclusive evidence that alpha dominant and rare alpha individuals differ on factors contributing to learning—certainly not on intelligence (11). The *Ss* were chosen with above average IQ's to maximize the probability of learning. In the preliminary work, *Ss* with considerably less dominant alpha rhythms were sometimes used. Although the borderline between drowsiness and sleep was more difficult to detect from the EEG with these *Ss*, material presented when delta activity was present was never recalled later.

Experimental and control groups were tested under slightly different conditions. The control group in the present study was given one unaided pretest followed by the multiple-choice posttest with no intervening training. The experimental group received an unaided pretest followed by training and then a second unaided posttest before taking the multiple-choice posttest. To what extent the additional unaided posttest effects the comparison between groups is unknown. However, it has generally been shown that additional practice such as this second test provided enhances learning (6). This again would increase the probability that the experimental group would show even more learning on the multiple-choice test than the control. This was not found to be the case in the Levels B through E associated with sleep.

Experimental *Ss* were given each item only once during training.[3] This was necessary to correlate an answer with a sleep level, for sleep levels varied throughout the night and had items been repeated, there would have been no assurance that they would have occurred in the same levels each time. Neither Stampfl (17), Hoyt (7), nor Coyne (2) found that varying the number of repetitions had any effect in their sleep-learning studies. None of these investigators found evidence of sleep-learning. The meaningfulness and simplicity of the test items tend to favor retention (6); tests showed that approximately 86% of previously unknown items could be retained by awake and alert *Ss* with a single repetition.

The above three limitations were accepted by the *Es* since in most cases they should tend to enhance any sleep-learning potential that might exist and thus disfavor the hypothesis which was tested. In spite of this, the evidence for learning was inconsequential as long as a rigid control on the measure of the depth of sleep was maintained.

Although it appears that learning during real sleep is not feasible, the practical utilization of the drowsy state for training is still open to speculation. The results in this study show that approximately 30% of the simple and highly organized material presented in the period just prior to sleep was recalled. Just how efficient learning in this state is has not been sufficiently evaluated experimentally. Anderson (1) found that the amount of recall decreases when training follows sleep closely. One must weigh the advantages of limited learning against the possible harmful effects from loss of sleep as well as against the time demanded to learn in the subnormal receptive state. It may be that in the drowsy state preceding sleep, the individual is more susceptible to suggestion; perhaps one's attitudes or habits can be modified during this presleep period when criticalness is minimized (8). Perhaps the future development of new and unknown techniques will permit someone to learn complex material while he sleeps, but for the present, sleep-learning is not the simple matter that some *Es* and commercial firms, which sell equipment for this purpose, would lead us to believe.

[3] Prior to the publication of this paper a second study (4) was completed in which material was presented to *Ss* between 16 and 82 times. No learning occurred so long as the material was presented in non-alpha levels.

Summary

Twenty-one *Ss* were selected on the basis of IQ (average or above) and an EEG showing a continuous alpha rhythm when they were awake and resting with their eyes closed. They were given a preinformation test of the sleep-training material and then permitted to sleep for 8 hr. No attempt was made to control the depth of sleep although EEG's were taken continuously from the occipital region. Ninety-six questions and answers were played to *Ss* once each at 5-min. intervals throughout the sleep period. The *Ss* were instructed to indicate immediately if they heard an answer. On awakening the next morning, *Ss* were given the questions and asked to write the answers. Following this they were given five alternative answers and asked to check the correct one.

EEG patterns occurring during stimulation were assigned to one of eight sleep levels. Both the percentage of immediate responses and the percentage of items recalled correctly decreased as the percentage of alpha frequencies decreased. Shortly after occipital alpha frequencies disappeared from the EEG, immediate responses and item recall also stopped. This was more evident when tested by recognition than by unaided recall. Approximately 5% of the items presented during non-alpha levels were responded to immediately, recalled later, or both. A majority of these items occurred when particular EEG patterns associated with arousal occurred; alternative explanations other than sleep-learning are offered for the few remaining items. Learning during real sleep is concluded to be impractical and probably impossible. The possibility of utilizing the drowsy state where material is retained is discussed.

References

1. Anderson, D. V. The effect of relaxation on the recall of nonsense syllables, words, and poetry. Unpublished doctor's dissertation, Univ. of California, Los Angeles, 1953.
2. Coyne, M. L. Some problems and parameters of sleep learning. Unpublished honors' thesis, Wesleyan Univ., 1953.
3. Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., & Hobart, G. Human brain potential during the onset of sleep. *J. Neurophysiol.*, 1938, **1**, 24–38.
4. Emmons, W. H., & Simon, C. W. The non-recall of material presented during sleep. Santa Monica, Calif.: The RAND Corp., Paper No. 619, 1955.
5. Gibbs, F. A., & Gibbs, E. L. *Atlas of electroencephalography*. Cambridge, Mass.: Lew. A. Cummings Co., 1941.
6. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 613–689.
7. Hoyt, W. G. The effect of learning of auditory material presented during sleep. Unpublished master's thesis, George Washington Univ., 1953.
8. LeShan, L. The breaking of a habit by suggestion during sleep. *J. abnorm. soc. Psychol.*, 1942, **37**, 406–408.
9. Lindsley, D. B. Electroencephalography. In. J. McV. Hunt (Ed.), *Personality and behavior disorders*, Vol. II. New York: Ronald, 1944. Pp. 1033–1103.
10. Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473–516.
11. Lindsley, D. B. Psychological phenomena and the electroencephalogram. *EEG clin. Neurophysiol.*, 1952, **4**, 443–456.
12. Loomis, A. L., Harvey, E. N., & Hobart, G. A. Cerebral states during sleep, as studied by human brain potentials. *J. exp. Psychol.*, 1937, **21**, 127–144.
13. Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 3 (Whole No. 142).
14. Moruzzi, G., & Magoun, H. W. Brain stem reticular formation and activation of the EEG. *EEG clin. Neurophysiol.*, 1949, **1**, 455–473.
15. Simon, C. W., & Emmons, W. H. Considerations for research in a sleep-learning program. Santa Monica, Calif.: The RAND Corp., Paper No. 565, 1954.
16. Simon, C. W., & Emmons, W. H. Learning during sleep? *Psychol. Bull.*, 1955, **52**, 328–342.
17. Stampfl, T. The effect of frequency of repetition on the retention of auditory material presented during sleep. Unpublished master's thesis, Loyola Univ., Chicago, 1953.

(Received April 1, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

THE PRECISION OF CHOICE REACTIONS[1]

E. C. POULTON

Applied Psychology Research Unit, Cambridge, England

In the early experiments on choice reactions, the appropriateness of a reaction was noted, but no attempt was made to record its precision (**5,** pp. 298–339). The current information-theory model of choice reactions has simply followed this tradition. Each choice reaction has been treated as identical with every other reaction of its particular order, regardless of its precision (**2**). This greatly reduces the use which can be made of models derived from reaction experiments, in the study of simple continuous skills such as tracking. For in these continuous tasks the error is a function of precision, as well as of speed of reaction. In tracking with an adequate view of the course ahead, precision is the more important concept. The error in terms of the average time lag behind the course is small compared with the error in terms of the average time off course (**3,** Table 3).

The present experiment was designed primarily to determine whether a one-choice reaction is made with greater precision than a two-choice reaction. A subsidiary aim was to determine whether the precision and reaction time (RT) of a two-choice reaction are altered by the insertion of 50% of catch tests (**5,** p. 308).

[1] The influence of Professor Sir Frederic Bartlett, F.R.S., is clearly shown in this paper. Mr. E. R. F. W. Crossman advised me in the interpretation of the graphic records. Miss V. R. Cane gave advice on statistical methods. Mr. C. B. Gibbs loaned the apparatus which produced the signals. The *S*s were supplied by the Royal Navy. Financial support from the British Medical Research Council is also gratefully acknowledged.

METHOD

A paper tape moving at 3.9 cm. per sec. carried (*a*) a center line, (*b*) a target line 1.0 in. to the right, with an additional line on each side of it .125 in. away, and (*c*) a target line 2.0 in. to the left of the center line, with an additional line on each side .25 in. away. The *S* held a ball-point pen, which had to be moved against a transparent bar lying at right angles to the direction of movement of the tape. There was a bell at head height on each side of him at a distance of about 6.0 ft. Whenever he heard the bell on his right ring, he had to make a quick 1-in. out-and-back movement to the right from the center line, aiming to change direction on the right-hand target line. Whenever he heard the bell on his left ring, he had to make a corresponding 2-in. movement to the left. The lines on each side of the target lines gave the permitted limits of precision for the positions of the changes in direction.

As the bell on one side rang, the pen drawing the inner of the two permitted limits of precision on that side was deflected. The pen was screened from *S*. It was mounted .5 in. away from the transparent bar along which his pen traveled. At the speed of movement of the paper tape, the mark made by the deflection took .33 sec. to reach the bar. His outward movement from the center had to cross the line drawn by this pen before the mark made by the deflection reached the bar. By looking at the paper tape, he could see at once if his reaction had met these criteria of time and precision. The time limit was applied in order to ensure that he reacted as soon as he could after the presentation of the signal. His state of preparedness at the time of presentation of the signal would thus be his state of preparedness as he reacted.

Each bell rang for about 1 sec. No warning signal was given before a bell started to ring, but the interval between the start of one ring and the start of the next was either 2.1 or 4.2 sec. Series of signals thus contained four components of combined time and direction. In the one-choice condition the order of the four components was fixed as follows: quick left, delayed left, delayed right, quick right, quick left. . . . In the two-choice condition the order of the components was random, except that every series contained at some point a group of four patterned

components, each of which had the same relationship to the previous component as in the one-choice condition. The patterned section was inserted into each two-choice series, in case the one-choice sequence was intrinsically easier or more difficult to respond to than the random sequences of the two-choice condition. The *S* was told that the order of the four components in the two-choice condition was quite random. In half the trials a bell would ring 2.1 sec. after the start of the previous bell. A bell would ring 4.2 sec. after the start of the previous bell, if it had not rung after 2.1 sec.

Each *S* received four combined practice and test periods of about 1 hr. each on different days in the same week. The half of the *S*s who comprised Group 12 spent the first two periods on the one-choice condition and the last two on the two-choice condition. The half who comprised Group 21 performed the two conditions in the reverse order. Two *S*s were tested each week, one from each group. The composition of the groups was equated as far as possible for intelligence and scholastic achievement.

Before performing the one-choice condition, *S* was taught the repeating pattern, and practiced responding to it. A similar amount of instruction and practice was given before the two-choice condition. It was designed to persuade *S* that the order of the components in the test series would be random. Eight test series were performed each day. The 16 test series of the two-choice condition were presented in random order to each *S*. Series contained from four to six practice signals, and then 12 test signals. Before each series *S* was told how many practice signals there would be. An increase in the speed of the paper tape as the last practice signal was presented indicated that the test signals were about to begin. Immediately after the completion of each series, the paper tape was examined by *S* and *E* together. Reactions which satisfied the combined criterion of time and precision were scored as correct, and *S* was encouraged to try even harder. He reacted altogether about 300 times under each condition.

Scoring.—A correct reaction (see Table 1) had to satisfy the combined criterion of time and precision. A reaction which started before the signal (premature) was scored like a normal reaction unless it was followed by a reaction at the proper time. A reaction which started toward the wrong target (start wrong) generally merged into a reaction toward the correct target, and only the latter was scored. But in both conditions about one-fifth of the wrong starts were not followed by an adequate correction. They thus correspond to errors in the conventional experiment on choice reactions. With about 40% of the double outward movements in both conditions, only the first part of the movement was counted. With the remainder, the two parts were scored as a single reaction.[3]

Median RT's are shown in Tables 1 and 3

[3] A more detailed discussion of the criteria used for scoring atypical reactions has been filed in the American Documentation Institute. Order Document No. 4702, remitting $1.25 for microfilm or $1.25 for photocopies.

TABLE 1

SUMMARY OF MAJOR RESULTS

Condition	Frequency of Categories of Reaction						Reaction Measures		
	Correct	Over-shoot	Under-shoot	Prema-ture	Start Wrong	Double Out. Movement	Mean Error (Mm.)	Median RT (Sec.)	Mean Speed (Cm./Sec.)
One-choice									
Group 12	50%	20%	8%	7%	.6%	33%	3.7	.14	35
Group 21	50%	22%	10%	12%	1.2%	16%	3.6	.14	35
Two-choice									
Group 12	27%	22%	17%	0%	9%	15%	4.1	.19	38
Group 21	9%*	27%	18%	.1%	9%	12%	4.9*	.26*	41
P (One-choice vs. two-choice)									
Group 12	.02	—		—	.001	.05	—	.01	—
Group 21	.001	—	.04	.05	.01	—	.01	.02	—
Both Groups	.001	—	.02	.01	.001	.05	.02	.01	.05

Note.—Only *P* values of .05 or better are shown.
* Two-choice Group 12 vs. Group 21 $P = .01$.

because the distributions were skewed (4). The significances of the differences in RT were calculated from the medians of the individual *Ss*, using *t*. The speed of movement was determined after excluding the initial acceleration and final retardation. Two-tailed tests of significance have always been used.

Subjects.—The 16 *Ss* were all enlisted men in the British Royal Navy, aged between 18 and 24 yr.

Results

Table 1 shows the differences between the one-choice and two-choice conditions for both groups of *Ss*. No significant difference was found between reactions after 2.1 sec. and reactions after 4.2 sec. The results for the two-time intervals have therefore been pooled. The only significant difference between the patterned and random parts of the two-choice condition was in the number of reactions which started in the wrong direction. With both groups of *Ss* the patterned parts gave more wrong starts ($p = .02$ for the two groups combined). As there were no other differences between the two parts, the results have been pooled. It was not possible to ask *Ss* directly whether they realized that a patterned part was present in every two-choice series, in case this fact was transmitted to subsequent *Ss*. But *E* formed the impression that if any of them were aware of it, the awareness was by no means clear enough to be verbalized. The serial position of the test item also made no difference.

TABLE 2

Distributions of RT's

RT (Sec.)	One-Choice*	Two-Choice	
	Groups 12 and 21	Group 12	Group 21
Shorter than .10	14%	2%	—
.10–.14	46%	19%	4%
.15–.19	34%	36%	21%
.20–.24	4%	20%	23%
.25–.29	2%	16%	20%
.30 and longer		7%	32%

* The distributions for the two groups did not differ significantly under this condition.

No reaction was successful if its RT was as long as .22 sec. In the two-choice condition no reaction was successful if its speed of movement was less than 17 cm./sec. There was a significant negative association in the two-choice condition between overshoots and undershoots, when the performances of the different *Ss* were compared (tau = − .60).

The distributions of RT's for the one-choice and two-choice conditions are shown in Table 2. The means

TABLE 3

Effects of Length of Movement

Condition	Mean Error (Mm.)			Median RT (Sec.)			Mean Speed of Movement (Cm./Sec.)		
	2-in. Movt. To Left	1-in. Movt. To Right	*P* (2 in. vs. 1 in.)	2-in. Movt. To Left	1-in. Movt. To Right	*P* (2 in. vs. 1 in.)	2-in. Movt. To Left	1-in. Movt. To Right	*P* (2 in. vs. 1 in.)
One-Choice* Groups 12 and 21	4.0	3.3	.02	.14	.14	—	40	30	.01
Two-Choice Group 12	4.9	3.4	.001	.18	.20	—	46	30	.01
Two-Choice Group 21	5.4	4.3	.05	.24	.27	—} .05	45	36	.05

Note.—Only *P* values of .05 or better are shown.
* The two groups were not significantly different on any measure under this condition.

were: one-choice .08 sec., two-choice Group 12 .20 sec., two-choice Group 21.27 sec. The RT's of the wrong starts were taken as the times until *S*'s pen began to move in the correct direction. However, the RT's of the initial incorrect movements did not differ significantly from the distributions shown here.

The reaction measures for right and left movements are shown in Table 3. There was no significant difference between right and left in the frequency of any of the categories of reaction listed in Table 1.

Discussion

Number of choices and precision.—The one-choice reaction has been described as a "prepared reflex" (**5**, p. 305). In the two-choice situation *S* cannot be as ready as this. For he has to be prepared to react in both directions, and also to inhibit whichever of the two prepared reactions is not required. The results show that this reduced readiness is reflected in reduced precision, as well as in a longer RT. It may be the readiness to inhibit which is reflected in the increased tendency to undershoot (Table 1).

Catch tests under two-choice conditions.—In the two-choice condition a signal could be expected after 2.1 sec. in only 50% of the trials; whereas if no signal was presented after 2.1 sec., a signal was certain 2.1 sec. later (see Method). It was predicted that the 50% of catch tests at 2.1 sec. would increase the RT at this time interval, and possibly also reduce precision, since in one-choice experiments the insertion of a large number of catch tests is known to increase the RT (**5**, p. 308). The finding that there was no difference between the two time intervals in the two-choice condition failed to confirm this prediction. Failure was not due simply to the greater difficulty of reacting after the longer time interval. For there was also no difference between the two time intervals in the one-choice condition in which there was no uncertainty in the times of presentation of the signals.

The result is consistent with Hick's view that "no stimulus" is treated "as if it had the same probability as any particular stimulus" (**2**, p. 13). For in the two-choice condition *S* had to wait for the signal before reacting, or he ran the risk of starting in the wrong direction. And the inevitable uncertainty in the exact time of presentation of each signal (whether after 2.1 or 4.2 sec.) was equivalent to one or more catches per signal. Hence, in Hick's view, the additional catches introduced by *E* should have made no difference.

The RT model for tracking in one dimension is thus two-choice, not three-choice as previously believed (**3**, p. 405). The two choices are to increase or decrease speed, or whatever differential of position with respect to time *S* is reacting to. Not to change in speed does not need to be counted as a third choice, since it simply corresponds to the catch tests of the present experiment.

Very short two-choice RT's.—The RT distributions in the one-choice condition (Table 2), and of Group 21 in the two-choice condition, are consistent with previous results (**5**, pp. 321–325 and 332–333). But the RT's of Group 12 in the two-choice condition are about .05 sec. shorter than usual, although there were only 2.1% of errors in the conventional sense (uncorrected wrong starts). This result appears to have been due to the incentive of immediate knowledge of results, following practice and a high level of aspiration in the one-choice condition. A reduction in the one-choice RT by incentives has been demonstrated previously (**5**, pp. 322–323).

Interaction between lengths of movement.—Table 3 shows that the mean error of the 2-in. movements to the left was not double that of the 1-in. movements to the right, although the permitted limits of error for a correct reaction were twice as large. Thus the 2-in. movements satisfied the criterion of precision significantly more often than the 1-in. movements ($P = .01$ for the two conditions, combined). However,

the speed of the 2-in. movements was not twice that of the 1-in. movements (although in the two-choice condition the increased time taken by the 2-in. movements was partly compensated for by a shorter median RT). Thus in both conditions the two sizes of movement were correct about equally often by the combined criterion of time and precision.

These findings for precision and speed of movement are different from those of Fitts for reciprocal tapping (1). Fitts found that *S* adapted his rate of responding pretty well to the permitted limits of precision. However, in Fitts' experiment *S* only made one size of movement in each experimental series, whereas in the present experiment two sizes of movement alternated in the same series.

Summary

It was desired to determine whether a one-choice reaction is made with greater precision than a two-choice reaction. A subsidiary aim was to determine whether a two-choice reaction is affected by the insertion of 50% of catch tests.

An auditory signal was presented either to the right or to the left of *S*, and he had to react in the corresponding direction with a 1-in. movement to the right or a 2-in. movement to the left. A criterion that combined time and precision had to be met. In the one-choice condition he knew in advance the direction and time of presentation of each signal, since the sequence of signals was arranged in a simple pattern that was learned before the condition was performed. In the two-choice condition he was told in advance that there were only two possible times of presentation, and only two possible directions of the signal, but that the order was random.

It was found that the one-choice condition was performed with greater precision than the two-choice condition. There was no difference in the two-choice condition between the two possible times of presentation of the signal. Thus 50% of catch tests had no effect in this condition. When the two-choice condition was performed after the one-choice condition, RT's were about .05 sec. shorter than those usually reported. There was evidence of an interaction between the 1-in. and 2-in. movements.

References

1. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47**, 381–391.
2. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.
3. Poulton, E. C. Eye-hand span in simple serial tasks. *J. exp. Psychol.*, 1954, **47**, 403–410.
4. Stevens, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.
5. Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received February 26, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

EFFECT OF LONG-TERM PRACTICE AND TIME-ON-TARGET INFORMATION FEEDBACK ON A COMPLEX TRACKING TASK [1]

E. JAMES ARCHER, GEORGE W. KENT,[2] AND F. A. MOTE

University of Wisconsin

Except for some very early experiments by Bryan and Harter (**4**) and Swift (**9**), there have been practically no long-term learning studies; and aside from one study (**2**) there is a complete lack of research examining on-target behavior in pursuit tracking. Most studies measuring performance in some tracking task have taken total cumulative clock time as a measure of goodness of performance. For example, in rotary pursuit experiments a circuit is so devised as to run a clock when *S* is in contact with the target and not operate the clock when *S* is not in contact. If a trial is of given duration, the determination of goodness of performance is easily expressed in terms of percentage of possible time on target. However, an important aspect of tracking behavior is lost with such a recording procedure. Good tracking behavior not only includes being on target for a high percentage of the time but also includes the acquisition of the skill of staying on target for long intervals continuously. A cumulative clock score does not differentiate between being on target once for 1 sec. continuously and 10 times for .1 sec. although the first tracking behavior is superior to the second. The most obvious way to obtain records of time continuously on target is with an ink-writing polygraph. It is also obvious that the reading of such records is a very tedious and time-consuming operation. The present study employed electronic sequential timers which counted the frequency of durations of time continuously on target.

The results of a study by Reynolds and Adams (**8**) suggest that good performance, when defined as long intervals continuously on target, might be learned through the use of some delayed feedback signal. In their study, *S* was given a click as a signal after he had been on target continuously for a stated duration. The required continuous on-target duration was systematically varied and an optimal duration was found. Of greatest interest was the result of a nonreinforced transfer test. When the click was no longer presented following a .5-sec. interval, the group which had been so reinforced during the training period continued to perform at a level superior to a control group that had never received a click. A variation of this procedure was used in the present experiment.

The present experiment was undertaken to study the change in complex tracking performance over a relatively long training period and to examine the effect of a delayed information-feedback signal upon time continuously on target. We would predict that as *S* improves in total time on target he does so not by simply in-

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-23294, monitored by the Armament Systems Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lowry Air Force Base, Denver, Colorado. Reproduction in whole or in part permitted for any purpose of the United States Government.

[2] Now at Bridgewater College, Bridgewater, Virginia.

creasing the number of hits of all durations but rather by increasing the frequency of long duration hits at the expense of shorter hits. In the light of the Reynolds and Adams study (8) we would expect superior performance by the reinforced group even though the on-target signal is delayed. Furthermore, if the results of Reynolds and Adams can be generalized to considerably more complex tracking tasks, the reinforced group should continue to do better than the control group when both are not reinforced.

Procedure

Subjects.—Thirteen paid, male University of Wisconsin students served in this experiment for 50 sessions.[1] None of the *Ss* had ever operated the apparatus although all had served in other psychological experiments. On the first day all *Ss* were read detailed instructions on the effective operation of the gunsight, and on the basis of the first day's performance, as measured by total time on target, the *Ss* were assigned to one of two groups. One group, hereafter designated the Control Group, was given no additional instructions on the second and succeeding days. They were simply reminded of their previous days' performance and encouraged to try to do better. The other group, hereafter designated the Experimental Group, was told that they would receive a special signal if they stayed on target continuously for 1 sec. This special signal was a 1000-cps tone presented over a pair of earphones. The tone was clearly audible but not irritating. Unlike the Reynolds and Adams study (8), the tone would continue to be heard if *S* remained continuously on target after 1 sec. For example, if *S* were to get on target for 1.3 sec. continuously, he would hear a tone for .3 sec. The *Ss* in the Experimental Group were told that they were members of a special group and were requested not to discuss the special signal with anyone. They could talk about their scores but not about the signal. The secret was fairly well kept—some *Ss* in the Control Group felt they knew after about 25–30 days. Other members of the Control Group did not know until after the 40th day. Knowledge of the other group's conditions was felt to be critical only insofar as the Control Group might perform at a lower level because they felt they could not or should not do well since they were deprived of a certain amount of knowledge of results. All *Ss* had at least six sessions per week, almost always on different days. Sometimes, because an *S* left town for the week end, it was necessary to have two sessions a day to keep him on schedule. In no case was an *S* permitted to have three sessions a day. If two sessions were run on a single day, they were separated in time by at least 6 hr. Starting with Session 37 all *Ss* had one and only one session per day every day including Saturday and Sunday. The entire experiment required 66 days to complete.

Apparatus and task.—A Mast Development Company modification of the Pedestal Sight Manipulation Test (PSMT) (6) was used in this study. This device has three major components (*a*) a flight generator unit which projects an image of the attacking fighter plane onto (*b*) a cylindrical screen 60 in. high and 93 in. wide, and (*c*) a pedestal sighting station. The latter was operated by *S* to track and frame the apparent attacking planes. The pedestal sight is mounted on top of the flight generator with its vertical axis coincident with the center of curvature of the screen. The radius of curvature of the screen is 5 ft. The pedestal sight projects a pattern of dots onto the screen. There is a center dot which is to be kept on the nose of the attacking plane by rotating the sight about its horizontal and vertical axes. This aspect of the sighting reticle enabled *S* to track in both azimuth (A) and elevation (E). In addition to the center dot there is a ring of dots, the diameter of which can be adjusted by *S* so as to match the wingspan of the attacking planes. This variation in diameter of the reticle was obtained by rotation of a vertical disc control held in *S*'s right hand and enabled him to track the target in the third dimension of range (R). The *S* operated the pedestal sight by kneeling in front of it and steadying the sight with his forehead while he viewed the target through an aperture in the sight. All *Ss* wore a pair of earphones during each session. The same tone which signaled the Experimental Group after a 1-sec. hit was given to all *Ss* about 5 sec. before the start of each of the four trials given during a session. This was a ready signal warning *S* to assume the proper operating position and to stand by for the first attacking plane. The *Ss* were instructed to stand up and relax during the 1-min. intertrial rest interval.

The sequence of attacking planes was controlled by a programming unit in the flight generator. There were four trials of eight attacks during each session which lasted about 15 min.

[1] Initially 14 *Ss* were started in the experiment, 7 in each of the two groups. After 14 days of training, however, one of the *Ss* in the Experimental Group became ill and had to drop out of the experiment.

In all there were 16 different patterns of attacks at approximately the same speed. During a trial, attacks would start from alternate sides of the screen and move on a curving pursuit course to the opposite side in about 7 sec. There was a 5-sec. delay between successive attacks. At the start of an attack the wingspan of the plane would be small and would appear to increase as the plane closed the range and pressed the attack. The *S* therefore always had to gradually increase the size of the ring of dots. At the end of the eighth attack in a series, a red signal lamp would light, indicating the start of the 1-min. intertrial rest period.

Following the session-over signal, *S* left the darkened room in which he operated the sight and entered the hallway which contained *E*'s recording apparatus. At this time, irrespective of his status (Experimental or Control Group) *S* was informed of his performance in terms of all of the response measures which were recorded. Because it was feared that *S* might become bored with such a long-term study and lose interest in doing as well as possible every day, intersession knowledge of results was maximized. To provide social motivation to do well, a graphic record of each *S*'s day-to-day performance in percentage of total time on target (AER), i.e., simultaneous tracking in all three dimensions of azimuth, elevation, and range, was posted in the hallway next to *E*'s recording consoles. No indication was made on the graphic records which *Ss* were in the Control Group and which were in the Experimental Group. (The records were arranged alphabetically.) Because some of the *Ss* knew each other quite well (members of the same fraternity) additional motivation was provided by a standing monetary bonus for the best performance for the session. Actually there were two such bonuses, of 85 cents each, one for the highest percentage of total time on target and the other for the best score for time continuously on target. Motivation was high and competition was keen.

Recording apparatus.—The recording console provided with the Mast PSMT has Standard Electric Timer clocks which measured the total time *S* correctly tracked in each of the three component dimensions of azimuth (A), elevation (E), and range (R). In addition, another clock provided a measure of the time *S* simultaneously tracked in both azimuth and elevation (AE). A fifth clock indicated the total time *S* correctly tracked in all three dimensions simultaneously (AER). In order to score on any of these dimensions, *S* had to have a displacement error of no more than 5 mils in visual angle (about ¾-in. tolerance on the screen).

A relay was placed in parallel with the AER clock clutch to provide a signal to the Tracking-time Analyzer. This last-named device has a bank of 12 thyratron-condenser-resistor units which fire their associated counters sequentially depending upon the duration of the signal coming from the AER circuit. One counter recorded the number of times *S* got on target for a duration of at least .05 sec. If *S* stayed on target for .10 sec., the second counter would record. The succeeding counters had minimum durations of .20, .30, .40, .60, .80, 1.00, 1.20, 1.60, and 2.00 sec. To illustrate, if an *S* stayed on target for as long as 1.30 sec., all counters from .05 to 1.20 sec. would fire. Through appropriate computations a frequency distribution of hits of various durations was obtained. A record was kept of this distribution after each of the four trials given during a session. A similar record was made of clock scores. In addition, another relay was placed in parallel with the 1.0-sec. counter to provide the special signal to the Experimental Group.

Results

Total Time on Target

AER tracking performance.—The total-time tracking performance of major interest is that shown by *Ss* in all three of the dimensions simultaneously, i.e., AER. Figure 1 shows the increase in this tracking performance for both groups for the entire experiment. The means for the curves are based on seven *Ss* for the Control Group and six *Ss* for the Experimental Group. On the first day both of the groups were treated alike—no tone was given for hits of 1-sec. duration. For Sessions 2 through 40 the Experimental Groups received feedback and the Control Group did not.[4] As can be seen, there is a sharp initial rise for the first three days for both groups and then tracking performance re-

[4] Unknown to the *Es*, until after the first few weeks, some of the Control Group *Ss* detected an indication of a long hit. Although the audio-oscillator, which provided the on-target signal to the experimental *Ss*, was turned off when a Control *S* was being run, the 1-sec. duration relay still closed and the "cross-talk" static in the headset was eliminated. Only a few of the Control *Ss* discovered this relationship and their performance was no better than the other Control *Ss*. This apparatus difficulty was corrected.

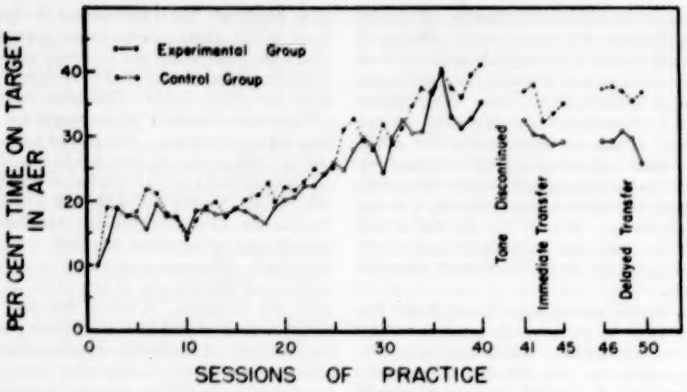

FIG. 1. Percentage time on target in AER as a function of stage of practice.

mains relatively unchanged, though highly variable, for about the next 18 sessions (approximately 18 days). Starting at about Session 20, performance begins to increase, attaining a maximum at about Session 40. An analysis of variance was used to test for the differences between the two groups and for the over-all change in performance. The results of this analysis are shown in Table 1. The difference between the means for the Control and Experimental Groups was not significant. These results indicate that the information feedback for the Experimental Group was ineffective in improving performance. In fact, the Control Group was numerically superior to the Experimental Group. The appropriate error term for evaluating individual differences and the effect of practice is the Pooled Residual/Groups. Both individual differences and practice were highly significant. To obtain a better evaluation of the practice effect, the sum of squares for Sessions was further analyzed with orthogonal polynomials (**5**). As may also be seen in Table 1 the linear, quadratic, and quartic components were highly significant. These results suggest that a fourth-order polynomial would provide a close approximation of the acquisition function. Apparently the initial rise, leveling until Session 20, the subsequent rapid increase, and second leveling at about Session 40 are all reliable phenomena.

TABLE 1

ANALYSIS OF VARIANCE OF AER SCORES FOR SESSIONS 1–40

Source of Variation	*df*	Mean Square	*F*
Groups	1	2,336.52	
Subjects/groups	11	3,690.77	22.97**
Sessions			
Linear	1	120,740.14	751.37**
Quadratic	1	6,588.33	41.00**
Cubic	1	135.77	
Quartic	1	2,929.88	18.23**
Residual from estimation	35	181.08	1.12
Sessions × groups	39	117.40	
Pooled residual/groups	422*	160.69	
Total	512		

* The Pooled Residual/Groups term was reduced by 7 *df* because it was necessary to make seven estimations of missing data. A malfunction in an amplifier of the range circuit gave perfect ranging scores and therefore inflated the AER scores by an unknown amount. These scores were discarded.
** $P < .001$.

Starting with Session 41 and thereafter, no special reinforcement was provided to the Experimental Group for hits in excess of 1-sec. duration. For the first five days in this transfer

situation the Experimental Group performed at a level numerically inferior to the Control Group. The difference between the means for the two groups, however, was not significant, an F of 4.84, for 1 and 11 df would be needed at the .05 level—an F of 3.22 was obtained. In Fig. 1 these sessions (41–45) are designated Immediate Transfer since they were given on the five days following the cessation of the tone signal. Since Session 37 all Ss had had one session per day.

The five trials designated Delayed Transfer in Fig. 1 (46–50) are delayed in the sense that 14 days' rest intervened between Sessions 45 and 46. These last five sessions also occurred one per day except for one S who had one day of rest between Sessions 49 and 50. Again, although the performance of the Experimental Group was numerically inferior to that of the Control Group, the difference between the means was not significant—an F of 4.34 was obtained when an F of 4.84 was needed for significance at the .05 level. For neither of the blocks of five transfer sessions was there a significant change as a function of practice.

Component tracking performance.—Although the increase in performance of the component skills of tracking in each of the dimensions A, E, R, and AE is intrinsically interesting, no findings were obtained which warrant any extensive discussion in this paper. Since the Tracking-time Analyzer was used only in the analysis of AER performance, only a record of clock scores was obtained for each of the component skills. For the first 40 sessions all components showed a significant increase. The F ratios for the four components for Sessions were 45.88, 41.44, 4.98, and 76.54 for A, E, R, and AE, respectively. (For 39 and 427 df an F of 1.65 is needed for significance at the .01 level.) In no case was the difference between the means for the Control and Experimental Groups found to be significant.

Time Continuously on Target

Frequency distribution of hits.—The change in the distribution of hits of various durations was not entirely as predicted. The number of short hits did decrease and the number of long hits did increase, but the total number of hits first decreased and then increased. These changes for each of the recorded durations as a function of stage of practice are shown in Table 2. The data have been grouped into blocks of five sessions each to simplify presentation. The values which appear in the table are means for all 13 Ss and represent the average total

TABLE 2

MEAN FREQUENCY OF HITS OF VARIOUS DURATIONS AS A FUNCTION OF STAGE OF PRACTICE

Duration of Hit (Sec.)	Sessions of Practice									
	1–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50
0.05	295.2	199.7	161.1	153.4	134.1	125.6	125.4	134.5	133.9	147.5
0.10	222.2	216.8	189.3	202.5	207.1	210.8	221.1	234.5	221.0	234.0
0.20	97.1	109.2	101.4	100.8	110.6	121.8	129.9	128.8	132.4	131.8
0.30	43.2	54.0	56.7	57.3	58.9	60.9	66.2	70.3	65.3	68.5
0.40	48.2	63.0	63.9	69.6	82.2	105.2	119.0	123.8	117.2	121.8
0.60	14.3	25.5	25.8	28.3	40.3	49.9	58.7	69.8	64.1	62.7
0.80	4.8	10.9	12.6	15.5	22.2	32.5	32.4	38.5	30.8	32.2
1.00	2.5	4.8	7.4	9.3	12.7	20.3	26.5	29.0	25.3	24.6
1.20	1.0	2.2	3.7	6.2	10.1	16.5	19.1	23.9	18.7	20.2
1.60	0.2	0.7	0.8	1.8	3.8	5.3	7.8	11.8	10.5	8.5
2.00	0.0	0.2	0.3	0.7	1.5	3.1	5.3	5.9	5.5	6.6
Total	728.7	687.0	623.0	645.4	683.5	751.9	811.4	870.8	824.7	858.4

TABLE 3

EXTENDED ALEXANDER TREND ANALYSIS OF VARIANCE OF NUMBER OF HITS OF .05 SEC. TO .09 SEC. DURATION

Line	Source of Variation	Error Term	*df*	Mean Square	*F*
a	Group means	b	1	2,225.68	
b	Individual means	n	11	19,721.40	18.002***
c	Group slopes	d	1	4,676.57	
d	Individual slopes	n	11	4,846.68	4.424***
e	Group quadratics	f	1	1,047.55	
f	Individual quadratics	n	11	2,091.25	1.909
g	Group cubics	h	1	33.40	
h	Individual cubics	n	11	2,279.22	2.081*
i	Over-all slope	n	1	160,082.62	146.128***
j	Over-all quadratic	n	1	130,874.24	119.465***
k	Over-all cubic	n	1	17,355.42	15.842***
l	Over-all deviations from estimation	n	6	1,566.01	1.429
m	Group deviations from estimation	n	5	349.85	
n	Individual deviations from estimation		67	1,095.50	
o	Total		129		

* $P < .05$.
*** $P < .001$.

number of hits made for a given duration for five sessions, i.e., an average of 295.2 hits equal to or longer than .05 sec. but shorter than .10 sec. were made by each *S* during the first five sessions.

There are several features of these data of particular interest. The shortest of the recorded durations, .05 sec., showed a very marked decrease with practice, attaining a minimum after about 30 sessions. The number of such hits then increased slowly for the remainder of the experiment. The significance of these changes were tested with an extension of Alexander's test for trend (1). A summary of this analysis is presented in Table 3. As may be seen from lines *a*, *c*, *e*, *g*, and *m*, the groups did not differ with respect to means, slopes, or other higher components. The individuals, however, differed in both means and slopes (as would be expected) and possibly the between-individual cubic component term was significant. Of major interest, however, is the over-all effect of practice on the number of hits of .05-sec. duration. In view of the significance of the over-all slope, quadratic, and cubic components, and the absence of significance of the over-all deviations from estimation, it seems probable that a third-order polynomial would give an adequate fit for the means appearing in the first row of Table 2.

The number of hits of .10-sec. duration shows a similar change with practice except that a minimum value is attained after about 10 sessions. These data (the means appear in the second row of Table 2) were also tested as before. Again the groups did not differ in means, slopes, or higher-order components; individuals differed in means and slopes; and the over-all slope and quadratic terms were significant at between the .05 and .01 levels of confidence. Apparently the number of hits of .10-sec. duration also decrease and then increase, but the change is not as marked as for hits of .05-sec. duration.

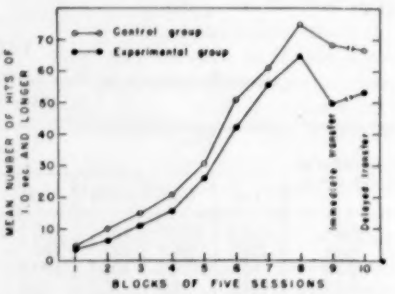

FIG. 2. Mean number of hits of 1.0 sec. and longer as a function of stage of practice. Experimental Group *S*s heard a continuous tone after they were on target for 1 sec. continuously. This tone was discontinued after Block 8. Fourteen days of rest was interpolated between Blocks 9 and 10.

For all the other durations the changes were less complex; the frequency of durations continuously on target longer than .10 sec. tended to increase with practice. If any differences between the Experimental and Control Groups were to be observed, it would be expected that the mean number of hits for durations in excess of 1 sec. should be greater for the Experimental Group since that group was specifically reinforced for such performance. These data, shown in Fig. 2, were also analyzed, using an extension of the Alexander trend test. As in all previous analyses the differences between the means, slopes, and higher-order terms for the two groups were not significant. As before, the differences between individual means were significant ($F = 27.886$, $P < .001$, for 11 and 45 *df*) as were the differences between individual slopes ($F = 6.293$, $P < .001$, for 11 and 45 *df*). Because of the addition of several learning functions (Rows 8 through 11 of Table 2 combined) the over-all, best-fitting function for these data is complex and probably meaningless (a fifth-order polynomial is indicated).

Although the two curves in Fig. 2 appear to diverge after the eighth block of five sessions, with the Experimental Group performing at a lower level, these differences were not significant. An analysis of variance of hits for 1 sec. and longer for the ninth and tenth stages of practice yielded no significant sources of variation.

Total time as a function of duration.—Another way of considering the effect of practice upon time continuously on target is to examine the change in percentage of total time accumulated at each of the recorded durations. Figure 3 shows the percentage of total time scored at each of the recorded durations for the first 40 sessions. Although the elimination of the special signal did not reduce the performance of the Experimental Group below that of the Control Group during the two transfer tests, it was decided not to include the data for these last two stages of practice. As indicated in Fig. 2 and in the subsequent trend test, the knowledge of a special test situation apparently altered the performance of both groups.

Although, as shown in Fig. 1, total time on target increased considerably during the first 40 sessions, the areas under the eight curves in Fig. 3 have been equated. The abscissa values for the curves are mid-points between the recorded durations of the Tracking-time Analyzer. The ordinates were determined by assuming a linear distribution of hits throughout the interval between two successive counters. The unequal interval sizes were equated by dividing the number of observed hits by a constant which was determined by dividing the interval size by .05. This corrected number of hits was then multiplied by the mid-point of the interval, and the recorded ordinate value is the ratio of this product to the total time on target.

There are several interesting features to these curves. The most obvious is the great reduction in the amount of total time accumulated at the shortest durations of time continuously on target as *S*s progress in practice from Block I to Block VIII. During the first five sessions of practice over 15% of the total time on target was scored by *S* hitting the scoring zone for durations less than .10 sec., whereas less than 3% of the total time scored during Sessions 36–40 was accumulated by such short hits. Another gross change in the curves is the increase in the area under the right-hand tails. This change reflects the increase in skill of staying on target for relatively longer durations. Because the curves have been plotted with equal areas, the differences have been minimized—if the ordinates indicated *total* time on target, the area under the uppermost curve (that for Block VIII of practice) would be about four times greater than shown.

A final point to note is the relative stability of the percentage of total time accumulated at the .50-sec. duration. Throughout the entire set of eight curves this value remains at about 5% and as the amount accumulated at the shorter durations decreases, the .50-sec. duration emerges as a second modal value.

Warm-up.—Several tests of warm-up decrement were made. The differences between the first trial on a given day and the last trial for the previous day were determined for both total AER and the longest duration of a hit. In neither case was there any

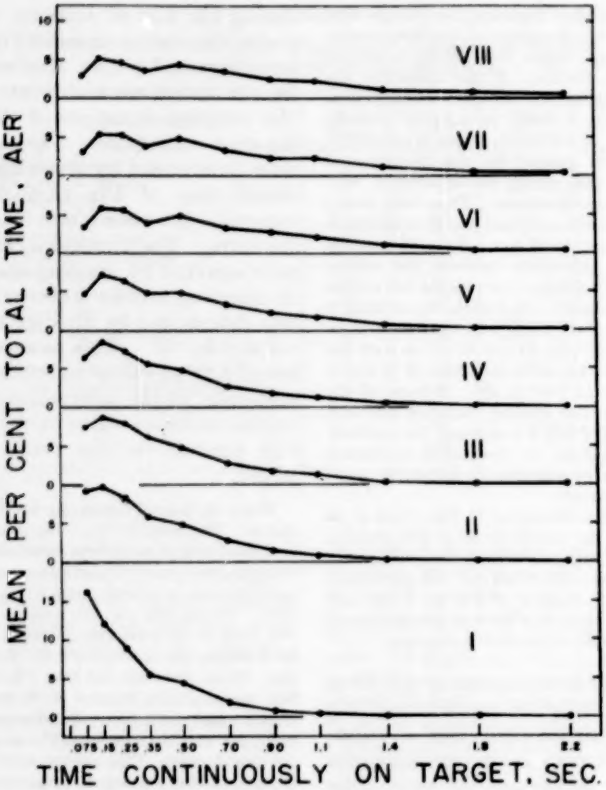

FIG. 3. Mean percentage of total time on target accumulated at various durations of time continuously on target as a function of stage of practice. The lowest curve (I) corresponds to the first five sessions of practice. The uppermost curve corresponds to Block VIII of practice (Sessions 36–40).

evidence for a warm-up decrement. Since the training period extended over a relatively long time, it was hoped that some systematic change in warm-up would be observed. Apparently the present task was so complex and the subsequent performance so variable as to conceal a systematic variation in warm-up. In addition to the previous tests, a similar comparison was made of the AE clock scores, since they did not include the great variability associated with range (R) tracking. As before there was no evidence for a warm-up effect.

DISCUSSION

There is little question about the effect of the delayed-information feedback on the performance of the Experimental Group. At least in this rather complex tracking task, delayed knowledge of results provided no facilitation in performance. These results may appear to be in disagreement with those of Reynolds and Adams (**8**) but it is believed that the contradiction is explicable.

In a rotary pursuit task, such as that used by Reynolds and Adams, *S* can easily see if he is *on* or *off* the target. If he is *off* he can also see how much and in which direction he must correct his tracking. In the present study the same decisions could be made as easily about the relation between the center dot of the reticle and the nose of the plane. However, such decisions were much more difficult for evaluating the accuracy of ranging (matching the diameter of the ring of dots and the wingspan of the plane). Since the accuracy of azimuth and elevation tracking was obvious to *S*, the tone signal's chief value was in testing the accuracy of ranging. However, since the results of this accuracy test were both delayed and dependent upon simultaneous correct tracking in the other two dimensions, and an attack only lasted for about 7 sec., the signal came too late to be of value to the Experimental Group. These conclusions are supported by the results of studies which introduced no delay between scoring in AER and the onset of a special signal (**3, 7**). In those studies performance was significantly improved if *S* received immediate knowledge of results, but as would be expected, the superior performance did not remain in a nonreinforced transfer test.

Although the drop in performance by the Experimental Group on both of the transfer tests was not significant statistically, the difference between the two groups is quite large—almost 8%. For the present, comment must be almost entirely conjecture, but starting with Session 37 the Experimental Group shows progressively poorer performance in percentage AER. As was pointed out previously, all *S*s had one session per day from Sessions 37 through 45. Since the Experimental Group knew they had the advantage of information feedback (that is, they thought it was an advantage), they would naturally expect to lose this advantage when they were all instructed to "get in step" and serve one and only one session per day. Since the *S*s were not naive to psychological experiments, their comments to the *E*s indicated that they suspected there would be a nonreinforced transfer situation. Many of the Experimental Group *S*s commented during the transfer tests that they simply couldn't score without being able to hear the tone. This conclusion was obviously erroneous since they did no better than the Control Group when they *did* have the tone. In short, the use of the tone not only did not improve performance during training but it became a psychological "crutch" which when removed led to diminishing performance.

The results of chief interest in this experiment are the changes in frequency of hits, changes in frequency of certain durations of hits, and finally the percentage of total time as a function of duration continuously on target. Table 2 suggests that *S*s smooth out their tracking performance, in terms of eliminating many of their almost saccadic sweeps of the target, faster than they develop the skill for staying on target for any long duration. The frequency of short-duration hits appears to decrease much more rapidly than the frequency of long-duration hits increases. The net effect is at first a decrease in total number of hits and then an increase. This relationship between changes in long and short hits is probably a function of task complexity. An easier task, like rotary pursuit, would just show an increase and then as *S* becomes very skilled, the number of hits would decrease to a minimum of one.

There is no question but that as *S* becomes better practiced in a complex tracking task such as a PSMT, he does not improve his score in total clock time simply by getting more hits at all durations of time continuously on target. Rather *S* improves his tracking performance by reducing the frequency of very short-duration hits and by markedly increasing the frequency of long-duration hits. In addition to this shift in the frequency distribution of hits, the total number also increases.

Summary

Thirteen paid, male students serves as *S*s in a 66-day experiment on the Mast Pedestal Sight Manipulation Test (PSMT). Seven *S*s (Control Group) received little or no intentional intra-

trial performance information feedback; the other six *Ss* (Experimental Group) heard a tone after they had been on target (scoring simultaneously on AER) for 1 sec. continuously. After 40 sessions of practice, the tone reinforcement was discontinued and all *Ss* were given five more sessions (days) of practice.

The Experimental Group showed a slight but statistically insignificant drop after the tone was eliminated. After two weeks' rest both groups were again given five sessions, one session per day. The Experimental Group was still numerically but not statistically inferior to the Control Group. The Control and Experimental Groups did not differ at any time during the study.

The major findings, with respect to time continuously on target, showed that as *S* improves in total time on target, the improvement is not a simple increase in the number of hits but rather a shift in the frequency distribution of durations of hits. The very short hits decrease in frequency and eventually account for but a small proportion of the total time on target, whereas the frequency of long-duration hits increases and accounts for a much larger percentage of total time on target.

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Ammons, R. B. Effect of distribution of practice on rotary pursuit "hits." *J. exp. Psychol.*, 1951, **41**, 17–22.
3. Bilodeau, E. A. Some effects of various degrees of supplemental information given at two levels of practice upon the acquisition of a complex motor skill. *Hum. Resour. Res. Cent. Res. Bull.*, 1952, No. 52–15.
4. Bryan, W. L., & Harter, N. Studies in the physiology and psychology of telegraphic language. *Psychol. Rev.*, 1897, **4**, 27–53.
5. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* New York: Hafner, 1953.
6. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviat. Psychol. Program Res. Rep. No. 4. Washington: Govt. Print. Off., 1947.
7. Morin, R. E., & Gagné, R. M. Pedestal Sight Manipulation Test performance as influenced by variations in type and amount of psychological feedback. *Hum. Resour. Res. Cent. Res. Note*, 1951, No. P&MS 51–7.
8. Reynolds, B., & Adams, J. A. Motor performance as a function of click reinforcement. *J. exp. Psychol.*, 1953, **45**, 315–320.
9. Swift, E. J. Studies in the psychology and physiology of learning. *Amer. J. Psychol.*, 1903, **14**, 201–251.

(Received February 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

FACILITATING EFFECTS OF INDUCED TENSION UPON THE PERCEPTION SPAN FOR DIGITS [1]

WILLIAM A. SHAW

University of Pennsylvania

Experiments dealing with the effects of induced muscular tension on mental tasks have differed methodologically and have yielded somewhat divergent results. In studies by Bills (**1**), Zartman and Cason (**6**), and Bills and Stauffacher (**2**) the weight pulled against to induce tension was more or less arbitrarily determined. Stauffacher (**5**) and Courts (**3**) varied the amount of tension as a function of *S*'s maximum pull. Facilitation was more clearly demonstrated in the latter two studies. Stauffacher (**5**) reports maximum facilitation when *S*s exerted one-half maximum pull; Courts (**3**) reports one-quarter maximum pull to be optimal. Conflicting data on the differential effect of induced tension on good and poor performers has been presented by Stauffacher (**5**) and by Bills and Stauffacher (**2**). The tasks used in previous studies have included learning nonsense-syllable lists (**1, 3, 5**), solving arithmetic problems (**2, 6**), and solving "detective story" problems (**2**). With the latter task, induced tension facilitated the solution of easy problems, but increased the time of solution of difficult problems.

The present experiment was designed to explore further the conditions for facilitation of performance by induced tension; specifically, to pursue further the matter of optimal tension as related to task difficulty and ability of *S*. The task used was the perception span for digits, a simpler task than those used in the previous studies.

[1] The author wishes to express his appreciation for a grant-in-aid from the Committee on the Advancement of Research, University of Pennsylvania; and to extend his thanks to Dr. Sherwin J. Klein and to Miss Annamarie Booz for their service in the conduct of the experiment.

METHOD

Subjects.—Thirty *S*s, 19 women and 11 men, took part in this study. All were undergraduates at the University of Pennsylvania who had taken one or more courses in psychology.

Design.—A factorial design was employed which permitted the evaluation of the variance attributable to tension, task difficulty, *S*s, and replication. Task difficulty was varied by varying the number of digits presented; 5, 6, 7, or 8 digits were presented in a given exposure. The induced tension used was zero and $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of each *S*'s maximum pull. Each *S* was exposed to two series of digits of the same length under each tension condition during the course of the experiment. That is, a total of 32 exposures of digits was viewed by each *S* during a single experimental session. The order of presentation of the series of digits and the weights (tension) was determined by a randomizing procedure prior to the running of *S*s. The order of conditions was the same for all *S*s.

The experiment was replicated two months later with the same *S*s. Different digits were used and a change was made in the procedure.

Apparatus.—The digits were exposed by means of a Dodge tachistoscope. The lights in the tachistoscope were switched on and off by a relay operated by a condenser timer. Commercial 40-w. light bulbs were used. The time exposure was adjusted so that several *S*s who were not in the experiment proper always perceived five digits correctly but not six digits.

The stability of the timer was not determined in the present experiment; previous calibration showed it to be highly stable. The plan of the experiment was such that any variation in the timer would be as likely to occur under one condition as under any other.

Tension was induced by means of a specially constructed apparatus. A weight was suspended from a lever in such a manner that it could be moved along a kilogram scale. A chain attached to the lever ran over a pulley and was attached to a handle on which *S* pulled. In-

duced tension was varied by varying the weight suspended from the lever. When the weight was lifted about 2 in. off the platform on which it rested, it tripped a microswitch which activated the timer controlling the tachistoscope. In the replication the microswitch turned on a signal light which was visible to *S*. The *S* himself then tripped the switch activating the timer.

Maximum tension was determined by attaching the handle on which *S* pulled to a spring scale and instructing *S* to "pull as hard as possible." In this way two instantaneous maxima were obtained; the average of these was used as the estimate of maximum pull.

The series of digits were typed on plain white cards. Natural sequences and numbers suggesting familiar dates were not used, nor was zero.

Procedure.—After *S*'s maximum pull had been determined, he was given the following instructions: "Your task during this experiment will be to repeat out loud series of digits which will be exposed in this apparatus. Pull that chair up to the apparatus so that you can sit comfortably while you look into this box here. [*E* points out opening.] Look into it now and you will see a cross. Do you see it? Before each exposure you must be sure you are looking at that cross. The digits will appear at that same point, and since the exposures will be rapid, your eyes must be fixated on that point or you will not see the digits when they appear. Now do you see this handle? When you pull on this handle the digits will appear where you now see the cross. During the experiment it will be necessary for you to pull this handle in order to make the digits appear. You will not know whether a slight or a hard pull will be required in any one case, so you must always use a slow gradual pull rather than a quick jerk. In pulling on the handle always keep your hand and arm above the top of the table and parallel to the table. Pull with your right hand and keep your left hand in your lap.

"I will always give you a ready signal before each series. Then when you are ready, pull slowly on the handle, keeping your eyes fixated on the cross. When the digits have appeared and quickly disappeared again, immediately repeat out loud the series you have just seen, and then relax until I give the ready signal again.

"Always be sure to repeat the digits accurately and in exactly the same order in which they appeared. If you cannot recall a certain digit in the series, say 'blank' for that digit and complete the series, e.g., upon failing to see the middle digit of the five digit number 43927, you would say '43 blank 27.'

"Now we will go through three practice trials to be sure you understand the procedure. The main things to keep in mind are: (1) In pulling on this handle with your right hand, use a slow gradual pull and keep your hand above the table and parallel to it. (2) Keep your left hand in your lap. (3) Before you pull on the handle be sure your eyes are fixated on the cross. (4) Be sure to repeat the series of numbers exactly as you see them. (5) Relax between each series of digits. Ready?"

All *S*s completed the task in one sitting. The experiment was repeated using the same *S*s and essentially the same procedure two months later. Changes were introduced to give *S* greater control over the apparatus. Instead of the lever tripping the switch controlling the tachistoscope, the lever tripped a switch which turned on a small signal lamp. When the lamp was lighted, *S*, at his own convenience, closed another switch with his unused hand to activate the tachistoscope. The tachistoscope could not be operated until the signal light was on. The series length and the order of presentation of series and weights remained the same. The actual digits in the various series were changed. The *S*s expressed greater satisfaction with this arrangement since they felt that they were always ready to respond when the exposure of the material occurred.

Results and Discussion

Of the numerous methods of scoring the digit span test—Guilford and Dallenbach (**4**) describe 27—the one of counting as correct only those digits correctly reproduced in the correct position was used. A summary of the analysis of variance of the results is presented in Table 1.

If presentations of 5, 6, 7, and 8 digits are thought of as forming a scale of increasing difficulty, performance on the series of digits of different lengths might be expected to show an effect due to different amounts of induced tension. The significant interaction between Tension and Number of Digits indicates that the various degrees of tension aroused by pulling on the weights affected performance on the series differently. Figure 1, a plot of this interaction, shows that the condition of one-fourth maximum pull facilitated the performance on all length series; the condition of one-

TABLE 1

ANALYSIS OF VARIANCE OF PERCEPTION SPAN SCORES

Source of Variation	*df*	Mean Square	*F*
Number of digits	3	122.81	102.34**
Tension	3	8.27	6.89**
Ss	29	20.69	17.24**
Sessions	1	20.01	16.68**
Tension × number of digits	9	4.74	3.95**
Tension × *Ss*	87	1.07	
Tension × sessions	3	.66	
Number of digits × *Ss*	87	2.12	1.77**
Number of digits × sessions	3	3.75	3.12*
Ss × sessions	29	1.45	
Number of digits × tension × *Ss*	261	1.03	
Number of digits × tension × sessions	9	.87	
Tension × *Ss* × session	87	1.12	
Number of digits × *Ss* × sessions	87	1.08	
Quadruple interaction	261	1.15	
Error†	960	1.20	
Total	1919		

* Significant between the .01 and .05 level.
** Significant at or beyond the .01 level.
† The error term employed throughout is the within-*S*, within-cell variance; that is, it is the variability of *S*'s performance under the same experimental conditions.

half maximum pull facilitated series of 6, 7, and 8 digits; and the condition of three-quarters maximum pull facilitated only 7- and 8-digit series. From these data it appears that while optimal tension is set up by one-fourth to one-half maximum pull, greater tension than this inhibits performance on the easier tasks but still facilitates performance on the more difficult tasks. The tension set up in the individual has, in general, a facilitating effect on the perception span for digits. This is also shown in Fig. 1 and confirms the results of Stauffacher (**5**) and of Courts (**3**).

The interaction between Subjects and Number of Digits is significant beyond the .01 level, whereas the interaction of Subjects and Tension and the triple interaction of Subjects, Tension, and Number of Digits are not significant. This indicates that although different *Ss* respond differently to the number of digits presented, they do not respond differently to induced tension nor is there a significant tendency for different *Ss* to be differently affected by certain weights during the presentation of certain numbers of digits. Inspection of the results for individual *Ss* indicates that for 14 of the 30 *Ss*, there was a monotonic relationship between performance and number of digits exposed, i.e., the more digits exposed the fewer reported correctly. For the other 16 *Ss*, one or two inversions in performance are found when comparisons are made between 5 and 6, 6 and 7, and 7 and 8 digits exposed. Presumably these inversions are the source of the significant interaction of Subjects and Number of Digits. On the basis of performance under the zero weight condition (no induced tension) the upper and lower 20% of the *Ss* were considered separately to see if good and poor performers responded differentially to induced tension. Fig. 2 gives the distribution of scores for

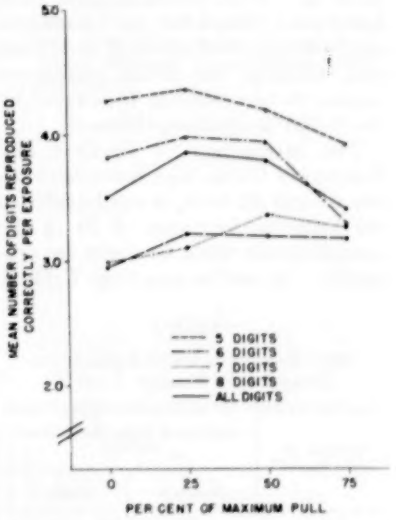

FIG. 1. Performance under different conditions of induced tension when different numbers of digits were exposed, and when all numbers of digits are considered together.

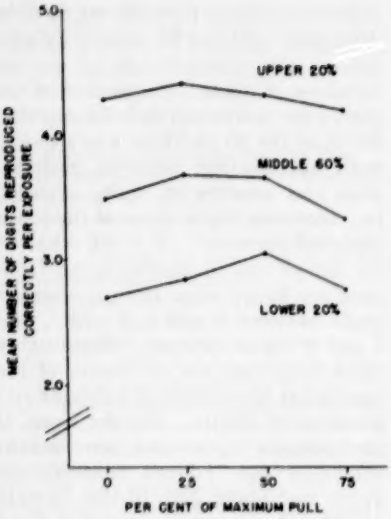

FIG. 2. Performance of *Ss* of three different levels of ability under different conditions of induced tension.

these *Ss*. Both extreme groups show facilitative effects for the conditions of one-fourth and one-half maximum pull, although the better performers appear to be optimally facilitated by the lighter weight condition.

The interaction of Sessions and Number of Digits, significant between the .05 and .01 level, is attributable to the lower performance of *Ss* in the second session when six digits are presented. As can be seen from Table 2, performance improved in the second session with respect to exposures of 5, 7, and 8 digits. The six-digit numbers used in the second session may have been more difficult than those used in the first session; no other interpretation was suggested by the data.

TABLE 2

MEAN NUMBER OF DIGITS REPRODUCED CORRECTLY IN SESSIONS I AND 2

Number of Digits Exposed	Number of Digits Reproduced Correctly	
	Session 1	Session 2
5	4.08	4.30
6	3.79	3.74
7	3.02	3.33
8	2.97	3.31

The variable Number of Digits is found to be significant at the .01 level. Is this merely due to one or more of the significant interactions in which Number of Digits is involved or can anything be said about the effect of Number of Digits per se? In Fig. 1 and in Table 2 it can be seen that for the different tension conditions and for Session I and Session II considered separately: Performance when five digits are exposed is superior to performance when six digits are exposed; performance on six digits is superior to performance for seven or eight digits; however, differences between exposures of seven and of eight digits are not consistent. The data for the interaction of Number of Digits and Subjects are also consistent with this finding. Therefore, the results for the single variable, number of digits exposed, show that performance declines as more digits are exposed; the smallest difference in performance as a function of number of digits exposed is that between seven and eight. This indicates that the assumption that number of digits exposed is a determiner of task difficulty is valid in this experiment.

The procedure in the second session differed from that of the first session only with respect to the means of activating the tachistoscope and the actual digits exposed. The difference between performance in Session I and Session II is highly significant. The *Ss*' performance was higher in the second session. The only interaction involving the Sessions variable, that of Number of Digits by Sessions, is attributable to lower performance in Session II with respect to exposures of six digits. Presumably, the superior performance in Session II was due to the change in procedure which permitted *Ss* to present the material to themselves when they were ready.

The type of task selected for the present study brings into focus more clearly than some of the tasks previously used, the fact that the facilitative effect of tension may occur very quickly, last only a relatively short time, and be fairly specific. For example, it cannot be argued that in the present experiment a general or diffuse tension due to attitude, to set, or to accumulative residual effects arising from pulling on the weights can be used to explain the results. Some generalized tension might have been present but its effects could only have reduced the difference between the effects of the different weight conditions which have the same effect in both sessions and on all *S*s. Nor does it seem likely that the facilitation shown can be explained as a result of increased attention since in the second session *S* controlled the exposure of the material and presumably attention was comparable under all tension conditions.

Summary

The effects of different amounts of induced muscle tension on task difficulty was investigated using the perception span for digits presented tachistoscopically. Optimal tension was found to vary as a function of task difficulty. One-fourth maximum pull facilitated performance on all length series of digits. Greater than one-fourth maximum pull facilitated the reproduction of the longer, more difficult, series more than the reproduction of the shorter, easier series. Induced tension facilitated the performance of both the good and the poor performers. The experiment was replicated two months later with the same *S*s. The effects of induced tension on performance did not differ significantly from the first experiment.

References

1. Bills, A. G. The influence of muscular tension on the efficiency of mental work. *Amer. J. Psychol.*, 1927, **38**, 227–251.
2. Bills, A. G., & Stauffacher, J. C. The influence of voluntarily induced tension on rational problem solving. *J. Psychol.*, 1937, **4**, 261–271.
3. Courts, F. A. Relations between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, **25**, 235–256.
4. Guilford, J. P., & Dallenbach, K. M. The determination of memory span by the method of constant stimuli. *Amer. J. Psychol.*, 1925, **36**, 621–628.
5. Stauffacher, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26–46.
6. Zartman, E. N., & Cason, H. The influence of an increase in muscular tension on mental efficiency. *J. exp. Psychol.*, 1934, **17**, 671–679.

(Received February 28, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

ACQUISITION OF MOTOR SKILL: IV. EFFECTS OF REPEATED PERIODS OF MASSED PRACTICE [1]

R. B. AMMONS AND LESLIE WILLIG [2]

University of Louisville

In the first article of this series (**3**), the senior author proposed a system for identifying and measuring certain characteristics of motor-skills performance. When, after continuous or nearly continuous practice, a rest period of more than a few minutes is introduced and then practice is resumed, the resulting performance curve often shows a sharp rise, followed by a relative decline, then a resumption of improvement. The first sharp rise is attributed to warming up after rest, and is measured by extrapolating backward the next section of the curve to the starting point of the period. The difference between this "corrected" initial level and the actual level of performance is the *warm-up decrement* (D_{wu}). The difference between the corrected initial level and the level that would have been reached had no rest period been introduced is used as an estimate of the amount of *temporary work decrement* (D_{Wt}, similar to Hull's I_R or reactive inhibition) present at the time of the last pre-rest trial. Finally, the amount by which this corrected initial level is less than the performance level of a group practicing under optimum conditions of distribution of practice is used as a measure of *permanent work decrement* (D_{Wp}, similar to Hull's ${}_sI_R$ or conditioned inhibition).

The same article (**3**) contained a review of the pertinent literature and a "miniature" theory consisting of assumptions and predictions. Certain of these assumptions and predictions have a direct reference to performance in the situation where numerous periods of continuous practice are given, and each period is followed by a rest of 5 min. or more: (*a*) It is assumed that amount of warm-up decrement (D_{wu}) will decrease as some function of the number of times practice is started. (*b*) It is predicted that amount of temporary work decrement (D_{Wt}) present at the end of the practice period will decrease as some function of the number of previous practice periods, at least after the first 8 min. of practice (**4**). This follows directly from the assumption that the decrease is a function of duration of prior practice. With a constant duration of practice period, the more periods there have been, the greater will be the total duration of prior practice. (*c*) It is assumed that amount of permanent work decrement (D_{Wp}) will increase as a joint function of the amount of temporary work decrement which has been present and the length of time during which it has been present [$D_{Wp} = f(D_{Wt} \times \text{time})$].

Numerous studies bearing on one or another of these statements have appeared since the publication of the original formulation and will be intro-

[1] The authors gratefully acknowledge a grant for 1951–52 from the Research Committee of the University of Louisville which made possible the completion of this work, and the able assistance of A. J. Eckles and R. G. Farr in collecting and analyzing a part of the data. Drs. C. H. Ammons and John Fink made many helpful suggestions during the preparation of the manuscript.

[2] Now at the State University of Iowa.

duced, as required, in the discussion of the present results.

The purpose of the present study was to determine the effects of repeated blocks of highly massed practice of rotary pursuit on amounts of warm-up decrement, temporary work decrement, and permanent work decrement developed.

METHOD

Subjects.—A total of 128 girls from Grades 11 and 12 served as *Ss*. The experiment was performed twice, each time with 16 *Ss* in each of four groups. As there was no appreciable difference between the two sets of results with respect either to statistical outcome or to appearance of curves, the data were combined, this giving four groups of 32. These four groups were matched as closely as possible for means and *SD*'s of scores during the first minute of practice, by dropping out six *Ss* in each group (see Table 1). This number had been arbitrarily set before the start of experimentation. Thus the final analysis was based on the scores of 104 *Ss*.

Apparatus.—The same four-unit pursuit rotor was used as in two previous studies (**5, 6**). Turntable surfaces were of black Vinylite, 10¾ in. in diameter. Brass targets were ¾ in. in diameter, set flush with the turntable surfaces, with their centers 3½ in. from the centers of the turntables. Each stylus was a wooden handle from which extended a hinged ¼-in. brass rod. This rod was 7½ in. in length, with an 80° bend 1¼ in. from the end. The hinged section rested on a silver tip with a weight of about .6 oz. During operation, the turntables turned at 60 rpm in a clockwise direction. When a steady contact between stylus tip and target was made, a 6-v. DC circuit was completed through relays and timers. The Standard Electric Timers recorded the duration of contact in .001-min. units. During operation, the turntables, targets, and stylus tips were cleaned every 10 min. with solvent.

TABLE 1

MEANS AND *SD*'S OF PERCENTAGE TIME ON TARGET DURING THE FIRST MINUTE OF PRACTICE

Group	*N*	Mean	*SD*
Continuous-continuous	26	5.3	5.2
Continuous-distributed	26	5.0	4.8
Distributed-distributed	26	5.4	4.9
Distributed-continuous	26	5.5	4.6

Design.—All *Ss* practiced for 90 min. in a training period, then practiced for 20 min. in a test period. There were two basic conditions: (*a*) continuous practice, during which *Ss* practiced continuously for 10-min. periods, with 20-min. rest periods between them; (*b*) distributed practice, during which *Ss* practiced for 1-min. periods, separated by 2-min. rest periods. These two conditions are referred to as 10–20, and 1–2, respectively. During the first 90 min., half of the *Ss* practiced in Cond. 10–20, half in Cond. 1–2. During the 20-min. test period, half of the *Ss* in each of these conditions were switched to the other condition. Thus there were four conditions, coded as indicated: (*a*) continuous-continuous (10–20, 10–20); (*b*) continuous-distributed (10–20, 1–2); (*c*) distributed-distributed (1–2, 1–2); and (*d*) distributed-continuous (1–2, 10–20).

This design was adopted for several reasons. It allowed for a direct comparison during the first 90 min. of the effects of high degrees of massing and distribution on temporary work decrement and permanent work decrement, while warm-up decrement could be calculated for the 10-20 groups. The switching of half the *Ss* in each group after 90 min. provided a necessary control, in that in a previous study (**6**) Ammons found that there was an interaction between decremental effects and distribution conditions. Reynolds (**19**) used a design similar to this in studying distribution-of-practice phenomena in eyelid conditioning, and Denny, Frisbey, and Weaver (**12**) as well as Ammons (**6**) in the case of rotary pursuit.

By keeping the work-rest ratio constant, the total time during which practice periods were given was very nearly the same for all groups. The 10-min. practice period provided for accumulation of a maximum of temporary work decrement (**4**), while the 2-min. and 20-min. rest periods gave time for recovery from this temporary work decrement (**4, 5**). Thus there were groups with highly massed and highly distributed practice. The highly distributed practice provided a control level with little or no D_{Wp}, D_{Wt}, and D_{ww}, while the highly massed practice produced a maximum of each of these.

Procedure.—The *Ss* were assigned randomly eight at a time, as available, to the four conditions. On entering the testing room, they were divided into two groups of four, and given chairs facing away from the apparatus, for use during rest periods. Each group of four was told how to hold the stylus and perform the task, and each *S* was assigned to a rotor. The general arrangements for the day were also explained: the groups were to practice and rest alternately

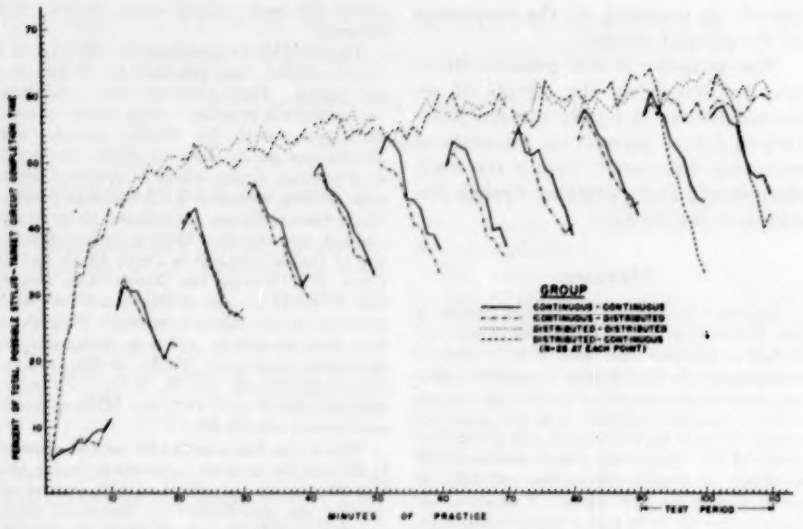

FIG. 1. Mean performance of the four practice groups.

as *E* indicated; during rest periods *Ss* could read magazines, but were not to watch the practice of members of the other group. After 50 min. of practice, lunch was provided, to be eaten during the next 20 min. of rest. *Ss* in Cond. 1–2 ate between 1-min. trials. A 5-min. recess was given after 60 min. of practice. Encouraging remarks were made after each 30 min. of practice.

When the 10–20, 1–2 and the 1–2, 10–20 groups were switched after the 90-min. training period, the rest periods between the training period and the test period were kept as nearly similar as possible to those in the training period, to avoid disturbance of quantity set. This meant that the 10–20, 1–2 *Ss* were given a rest of approximately 20 min., while half of the 1–2, 10–20 *Ss* rested 2 min. and half approximately 12 min. The 12-min. rest was necessary because two groups were run at the same time. Since one group started 10 min. of continuous practice after a 2-min. rest, the other group could not start until after approximately 12 min. of rest. Inspection revealed no appreciable difference between the scores of *Ss* receiving the 2-min. rest and those receiving the 12-min. rest.

RESULTS

Fig. 1 presents the mean performance by 1-min. periods for each of the four initially matched groups.[3] Proficiency increased greatly during the first 20 min. of practice, then much more gradually during the remainder of practice. Throughout, the massed practice groups made markedly lower scores, on the average, than the distributed groups, and in most cases showed a typical rapid rise at the start of sessions, followed by an equally rapid decline.

In order to make various calculations, it was necessary to obtain stable estimates of the performance level of the distributed groups at the start of each 10-min. period, the performance level of the massed groups at this point if no rest had been introduced, and the performance level of the massed groups corrected for D_{ww}. A straight line was fitted by the method

[3] Tables giving mean performance by trials and other data (see Footnotes 4, 5) have been deposited with the American Documentation Institute. Order Document No. 4739, remitting $1.25 for microfilm or $1.25 for photocopies.

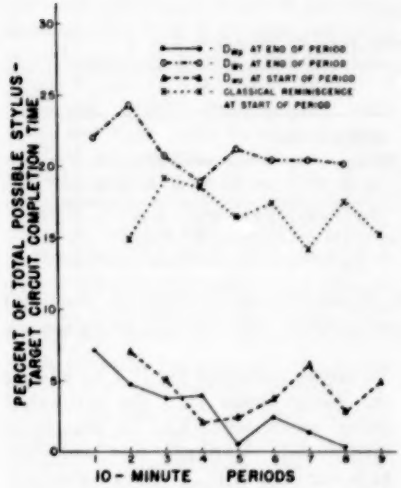

FIG. 2. Some derived measures of performance variables as affected by duration of prior practice.

of least squares[4] to the relatively decremental part of the curve for each 10-min. session of massed practice ($N = 52$), and then extrapolated backward to give the estimate of performance level on the first trial of the session, corrected for D_{wu}. Performance at Minutes 11, 21, etc. of distributed practice ($N = 52$) was estimated by least-squares[4] fitting of a straight line to the group mean scores for 10 trials at a time: the preceding five trials, Trial 11 (21, etc.), and the succeeding four trials. A line of the same slope was passed through the mean for the last trial in each 10-min. massed-practice session and extrapolated to give the estimated performance on Trials 11, 21, etc., had no rest been introduced. The measures of variables calculated in this way are presented in Fig. 2.[5] Methods and rationale of calculation are described in greater detail in an earlier article in this series (**3**).

Warm-up decrement (D_{wu}).—D_{wu} is the amount that performance is depressed below the level estimated by backward extrapolation for the first trial of each 10-min. session of massed practice. The mean D_{wu} for the two 10–20 groups combined, for Sessions 2 through 9, is given in Fig. 2. There is no indication of a systematic change in amount of D_{wu} associated with increasing proficiency or duration of practice, although there is considerable variability in amount from session to session.

A rough statistical evaluation of the over-all significance of changes in D_{wu} was made as follows. For each first trial (Sessions 2–9), the difference between the mean performance level corrected for D_{wu} and each *S*'s score on that trial was obtained. These numbers were arrayed in an 8-by-52 table, the 8 columns representing the 8 sessions (2–9) and the 52 rows representing *S*s. Variance was then analyzed into that associated with Sessions, *S*s, and residual, with 7, 51, and 357 *df*, respectively. Using the residual variance estimate as error term, the sessions and individuals variances were significant at beyond the 1% level (*F*'s of 4.17 and 23.19).

A further statistical test was made. The array of numbers in each of the columns (Trials 11, 21, etc.) was treated separately. A *t* test was made of the hypothesis that the true mean of the individual differences from the group mean score corrected for D_{wu} (i.e., the individual estimates of D_{wu}) was zero. The *t*'s so obtained are shown in Table 2. It can be seen that although there is a statistically significant variability (as evaluated by *F* test) in amount of D_{wu} from

[4] Equations are given in the document mentioned in footnote 3.

[5] Values are given in the document mentioned in footnote 3.

TABLE 2

SIGNIFICANCE OF DIFFERENCES FROM ZERO OF INDICES IN FIG. 2

Index	*t* Ratio									*F* Ratio
	10-Min. Periods									All Periods Together
	1	2	3	4	5	6	7	8	9	
D_{Wp} at end of period	4.19*	2.74	2.17	2.34	0.32	1.53	0.83	0.18		7.88**
D_{Wt} at end of period	13.10	13.86	12.03	11.11	11.46	12.29	12.07	12.24		3.15
D_{wu} at start of period		4.20	2.90	1.22	1.46	2.25	3.54	1.57	3.03	4.17

* *t* must reach 2.75, 2.04, and 1.70 for significance at 1%, 5%, and 10% levels, with 1 and 30 *df*. There were 1 and 51 *df* in this series of tests.
** *F* must reach 2.96 for significance at the 1% level, with 6 and 120 *df*. There were 7 and 357 *df* in each case.

session to session, and although five of the eight estimates of D_{wu} are significantly greater than zero (5% to 1% level), there is no apparent systematic change in D_{wu} with amount of practice.

Temporary work decrement (D_{Wt}, I_R).—D_{Wt} develops during continuous practice and dissipates during rests. This dissipation has been found to be complete or nearly so after 5 min. of rest (**4**). When rests of more than 5-min. duration are introduced, D_{Wt} on the last prerest trial can be measured as the difference on the first postrest trial between estimated level of performance if there had been no rest, and actual level corrected for D_{wu} by backward extrapolation of the subsequent curve. Most, if not all, of the observed sharp drop in performance during continuous practice periods in this study is due to cumulation of D_{Wt}, since the gains over rest bring performance nearly back to the level of the distributed practice groups. Average amounts of D_{Wt} on Trials 10, 20, etc., are shown for the massed practice groups in Fig. 2. D_{Wt} is large and may decrease slightly in amount as practice continues and proficiency increases.

An 8-by-52 table was made up, as in the case of D_{wu}, with individual values calculated by subtracting from S's score, corrected for D_{wu} by adding the group mean D_{wu}, the estimated group mean score had no rest been introduced. Variance was analyzed as in the case of D_{wu}. Mean amounts of D_{Wt} at the end of various sessions vary significantly ($F = 3.15$, with 7 and 357 *df*, with residual as error term). *T* tests of the null hypothesis at each point (values given in Table 2) indicate that significant amounts of D_{Wt} are present at the ends of all training-period sessions.

An interesting D_{Wt} phenomenon was noted. When *S*s switched from distributed practice during the first 90 min. to massed practice in the test periods (Group 1–2, 10–20), they showed a marked "sensitivity," developing much greater D_{Wt} than did either of the massed practice groups during the first 90 min. A test of the difference in trends (**18**) of the mean performance by the 1–2, 10–20 and the 10–20, 10–20 groups for Minutes 91–100 (first test period), showed them to be significantly different at beyond the 1% level of confidence ($F = 8.09$ with 9 and 450 *df*). The sensitivity seemed to be disappearing by the second test period (Minutes 101–110).

"Classical" reminiscence.—This is the difference between actual performance level on the first trial after

rest and estimated level of performance on this trial if no rest had been introduced. It is our D_{Wt} less D_{ww}. It can be seen in Fig. 2 that amount of "classical" reminiscence is quite variable, but that there is no consistent change as practice continues. The low point at 61 min. (start of seventh 10-min. period) is followed by a high point at Minute 71.

Permanent work decrement (D_{Wp} or $_{S}I_{R}$).—Tests for presence and amount of D_{Wp} have been made in two ways in the present study: (*a*) at the end of each practice period, and (*b*) during subsequent test periods (Minutes 91–110). At the start of each 10-min. practice period for the massed practice groups, it is possible to compare their performance, corrected for D_{ww}, with that of the distributed practice groups. Assuming on the basis of previous work (**4, 5**) that very little or no D_{Wt} is present for either the 1–2 or the 10–20 groups, any relative depression of the performance of the massed group is defined (**3**) as due to D_{Wp}, as developed by the end of the previous session. In Fig. 2, it can be seen that there is little D_{Wp} at the end of Period 1 (calculated at Minute 11) and the amount declines to nothing as practice continues.

Another 8 by 52 table was constructed, as for D_{ww} and D_{Wt}. Individual values were obtained by correcting each *S*'s score for D_{ww} by adding on the group mean D_{ww} for that trial and then subtracting the resulting score from the estimated optimum performance level on that trial (estimated level of 1–2 group, taken from least-squares line). Variance was analyzed into the same components as in the case of D_{ww} and D_{Wt}. Differences from session to session in D_{Wp} were highly significant ($F = 7.88$, with residual as error term, 7 and 357 *df*). *t* tests of the null hypothesis at each of the eight points (values given in Table 2) indicate that there is significant D_{Wp} at the start of practice, but not during the later stages.

The second method for testing for D_{Wp} involves switching half of the distributed-practice *Ss* to massed practice and half of the massed to distributed practice. If we assume approximately equal generalization of proficiency from each condition to the other, any D_{Wp} developed during the 90 min. of massed practice will persist, and result in lowered performance by the former massed-practice groups no matter what their current condition of practice.

Table 3 presents the results of the analysis of variance during the "test" period (Minutes 91–110). It should be noted that this analysis allows only a rough evaluation of the effects of the variables, since initially the groups were matched. Whereas the type of distribution currently employed has a relatively large and highly (statistically) significant effect, there is almost no detectable carry-over effect from the earlier conditions of distribution. Thus the two methods agree in show-

TABLE 3

ANALYSIS OF VARIANCE OF CODED SCORES FOR THE TWO TEST PERIODS

(MINUTES 91–100, AND 101–110)*

Source of variance	*df*	Mean Square
Distribution during first 90 min. (A)	1	4.32
Distribution during two test periods of last 20 min. (B)	1	3572.32**
Test periods (C)	1	64.69
A × B	1	200.07
A × C	1	46.17
B × C	1	62.47
A × B × C	1	44.31
Within groups	200	116.98

* A single score was obtained by adding *S*'s scores for one test period (30 20-sec. scores), then dividing this number by 100.

** Significant at 1% level with within-groups variance estimate as error term.

ing little or no permanent work decrement.

Discussion

As the present findings with respect to D_{wu}, D_{Wt}, and D_{Wp} bear out earlier predictions (3, 15, 16) only in part, it seems desirable to discuss available information pertinent to the disagreements. Some possible explanations will be suggested and briefly discussed.

Warm-up decrement (D_{wu}).—It was predicted earlier in this paper, on the basis of a prior review of the literature and attempted theoretical formulation (3), that after reaching an early maximum, amount of D_{wu} would decline in the case of the 10–20 groups in proportion to the number of rests and consequent re-startings of practice. This prediction was not borne out, since the amount of D_{wu} remained about the same. It may be that there were not enough startings of new sessions for *S*s to learn to overcome D_{wu}. More sessions could be introduced by having short rest periods (e.g., 30 sec.) between short trials (e.g., 15 sec.), with a single longer rest (e.g., 20 min.) after every 10-min. block of practice. It is exactly in this kind of situation where an increase in D_{wu} early in practice (9) and a decrease late in practice have been noted (1, 5).

Temporary work decrement (D_{Wt}).—Contrary to prediction, there is in the present study no appreciable decrease with increasing proficiency in the amount of D_{Wt} developed during a 10-min. period of continuous practice. Such decreases, where found, may be due to decreasing motivation (hence less tolerance of D_{Wt}) or to the distortion of units (and therefore of variability and of mean differences) as ceiling scores are approached. Kimble and Shatel (17) found a decrease of D_{Wt} to zero late in practice. We interpret this as an artifact of their method of measurement which did not correct for "forgetting" over the rests between blocks of trials or for more than a small amount of the D_{wu} present.

Several possible ways of accounting for the sensitivity to D_{Wt} shown during the test periods by the 1–2, 10–20 group seem plausible: (*a*) *S*s with relatively distributed practice (1–2) have had less opportunity to develop habits for responding efficiently when large amounts of reactive inhibition are generated; (*b*) *S*s with distributed practice rapidly lose motivation when they are given a long period of continuous practice; (*c*) the stimulus becomes increasingly different from that present during acquisition of the most adequate responses, and consequently fewer and fewer of these are elicited. This stimulus-change theory has been developed in some detail by Reynolds (19) to account for certain effects of massing and distribution in conditioning of eyelid closure. Interaction between acquisition and test condition, such as his theory would predict, was actually found for the rotary pursuit task by the senior author in earlier studies (5, 6), and has been noted by others (2, 13, 21, 23).

Permanent work decrement (D_{Wp}).—Contrary to prediction, little or no D_{Wp} ($_sI_R$, conditioned inhibition) was found in the present study by either the "backward extrapolation" or the "final test" method. Although the amount of D_{Wp} present at the end of each of the first four sessions is statistically significant, it is much smaller in amount than the 50% one would predict from Barch's survey (8), since it ranged from 18% to 33%. The decrease of D_{Wp} in the last four sessions to an average magnitude of only 6% is hardly what would be predicted of a habit ($_sI_R$) being learned.

Denny, Frisbey, and Weaver (12) suggest that when *S*s are changed from relatively massed to relatively distributed practice, conditioned inhibition (D_{Wp}, $_sI_R$) will be extinguished. Although this could help account for our negative results during the test period, it is not relevant to the decrease in D_{Wp} during the practice sessions, since there was no switching of distribution conditions. Kimble and Shatel's finding (17) of increasing amounts of D_{Wp} may well be due to their failure to correct for the D_{wu} and the "forgetting" which is clearly shown by the distributed-practice group.

It should further be noted that during the test period, the performance level of the 1–2, 10–20 group at the start of each 10-min. period is considerably higher than that of the 10–20, 10–20 group; and that the 1–2, 1–2 group shows higher performance than the 10–20, 1–2 group at 18 of the 20 points. These results may well be due to differential D_{ww}, with D_{ww} interacting with practice- and test-period distribution variables. Since no method has been developed for correcting for D_{ww} under such circumstances, these observations can only serve as a caution.

We should also be cautious about our estimate of optimum performance, based on the performance of the 1–2 groups. The actual performance of these groups is most certainly slightly depressed by D_{Wt} built up during each trial, even though this D_{Wt} dissipates completely over intertrial rests. There is also a possibility that the 1–2 groups develop some D_{W_p}. Both factors would somewhat depress our estimate of amount of D_{W_p} developed under 10–20 conditions, but not enough to account for smallness of the amounts or their decrease in later practice sessions.

Whatever the reason or reasons for it, the fact remains that apparently little or no D_{W_p} was developed by the massed-practice groups in the present study. Similar findings have been reported by Reynolds and Adams (**20**), Duncan (**13**), Adams and Reynolds (**2**), Starkweather and Duncan (**23**), Schucker, Stevens, and Ellis (**22**), Archer (**7**), and Reynolds and Bilodeau (**21**). This failure of D_{W_p} to develop is of considerable theoretical significance. Since massed and distributed practice took place over about the same total duration of time, the finding of approximately equal terminal proficiency lends additional support to a stimulus maturation theory of rotary pursuit learning (**10, 14**). It would seem that a contiguity theory would predict that D_{W_p} would develop (erroneous responses be fixated) during massed practice. Hence, the present findings are inconsistent with prediction from such a theory.

Since the systems of Hull (**15**), Kimble (**16**), and Ammons (**3**) all predict that D_{W_p} will develop during highly massed practice, we conclude with others (**2, 11, 21, 22, 23**) that theoretical formulations dealing with inhibition phenomena quite possibly stand in need of extension or revision. It may be that conditioned inhibition ($_sI_R$, D_{W_p}) is an unnecessary construct and that the phenomena referred to can be accounted for in other ways. A reconsideration of evidence for $_sI_R$ in conditioning might even be profitable in view of the present findings.

Summary

Four groups of *Ss* ($N = 26$ in each) practiced rotary pursuit for 110 min., 90 min. in training conditions and 20 min. in test conditions. The two basic conditions called for continuous practice (cycles of 10 min. work, 20 min. rest) and distributed practice (cycles of 1 min. work, 2 min. rest). The following four combinations of training and test conditions were used to make possible the measurement of warm-up decrement (D_{ww}), temporary work decrement (I_R), and permanent work decrement ($_sI_R$): continuous-continuous, continuous-distributed, distributed-continuous, and distributed-distributed. It was found that: (*a*) continuous practice led to poorer performance at all stages of practice; (*b*) proficiency increased rapidly for the first 20 min. of practice and more slowly thereafter; (*c*) warm-up decrement and classical reminiscence remained at about the same level throughout practice; (*d*) temporary work decrement did not decrease significantly as practice continued; and (*e*) there was little or no evidence of permanent work decrement. It was concluded that theories predicting the development of appreciable amounts of permanent work decrement paralleling temporary work decrement may need extension or revision, and that perhaps permanent work decrement ($_sI_R$) is an unnecessary construct.

References

1. Adams, J. A. Warm-up decrement in performance on the pursuit-rotor. *Amer. J. Psychol.*, 1952, **65**, 404–414.
2. Adams, J. A., & Reynolds, B. Effect of shift in distribution of practice conditions following interpolated rest. *J. exp. Psychol.*, 1954, **47**, 32–36.
3. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **45**, 263–281.

4. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
5. Ammons, R. B. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777–787.
6. Ammons, R. B. Relationship of motivation and method of testing to distribution of practice phenomena in rotary pursuit. *Quart. J. exp. Psychol.*, 1952, **4**, 155–164.
7. Archer, E. J. Postrest performance in motor learning as a function of prerest degree of distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47–51.
8. Barch, A. M. The relative contributions of I_R and $_sI_R$ to the total inhibitory potential. *Percept. & Mot. Skills Res. Exch.*, 1952, **4**, 27–31.
9. Barch, A. M. Warm-up in massed and distributed pursuit rotor performance. *J. exp. Psychol.*, 1954, **47**, 357–361.
10. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**, No. 1 (Whole No. 243).
11. Bilodeau, E. A. Rate recovery in a repetitive motor task as a function of successive rest periods. *J. exp. Psychol.*, 1954, **48**, 197–203.
12. Denny, M. R., Frisbey, N., & Weaver, J. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48–54.
13. Duncan, C. P. The effect of unequal amounts of practice on motor learning before and after rest. *J. exp. Psychol.*, 1951, **42**, 257–264.
14. Hilgard, E. R., & Smith, M. B. Distributed practice in motor learning: score changes within and between daily sessions. *J. exp. Psychol.*, 1942, **30**, 136–146.
15. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
16. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
17. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
18. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
19. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
20. Reynolds, B., & Adams, J. A. Effect of distribution and shift in distribution of practice within a single training session. *J. exp. Psychol.*, 1953, **46**, 137–145.
21. Reynolds, B., & Bilodeau, I. McD. Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.
22. Schucker, R. E., Stevens, L. B., & Ellis, D. S. A retest for conditioned inhibition in the alphabet-printing task. *J. exp. Psychol.*, 1953, **46**, 97–102.
23. Starkweather, J. A., & Duncan, C. P. A test for conditioned inhibition in motor learning. *J. exp. Psychol.*, 1954, **47**, 351–356.

(Received March 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

LEVEL OF CONDITIONING AND INTENSITY OF THE ADAPTATION STIMULUS

JANET A. TAYLOR

Northwestern University

In classical defense conditioning, the noxious UCS has typically been described not only as determining the characteristics of the response to be conditioned but also as being the (primary) source of motivation. In this latter role, the UCS might be considered to evoke an internal emotional response, as well as the overt UCR, the persistent effects of which, in turn, determine drive level.

The notion that the UCS has a motivational function has received experimental support in several studies (**5, 7, 8**) in which level of conditioning was shown to be positively related to intensity of the UCS. Of particular importance in this connection is a study by Spence (**7**). By employing a factorial design in which intensity of the UCS was shifted midway through the conditioning trials, Spence attempted to determine whether the differences in performance found with varying intensities of the UCS could be attributed to differences in learning or habit strength, in drive level, or in both of these. His results suggested that performance differences could be due, at least in part, to motivational properties of the UCS.

A second source of evidence that suggests that the UCS possesses emotional and motivational properties is found in an investigation by MacDonald (**4**). In the first part of her study, MacDonald found that the amplitude of the PGR in response to threat of shock decreased significantly after a series of 50 shocks to the finger had been experienced. Similarly, spontaneous blinking rate, taken as an index of emotional tension, was less after 50 puffs of air to the eye than after 5 puffs. Such a decrease in the response to noxious stimuli with repeated trials has also been shown in a number of other studies (**3, 6**). In the second part of her investigation, MacDonald attempted to determine whether the lowered amplitude of response to the noxious stimulation reflected a decrease in motivational level. The groups given the shock and the puff stimulations underwent a conditioning procedure, the UCS previously presented singly being paired with a neutral CS. The conditioning performance in the eyelid and finger withdrawal situation of the groups first given adaptation trials was found to be significantly lower than that of control groups given only conditioning trials. Thus the results were consistent with the hypothesis that the magnitude of the emotional response evoked by the UCS during conditioning and hence the amount of drive were less as the result of the adaptation trials.

A question arises as to whether the magnitude of the emotional response and hence the drive-evoking capacity of the UCS during conditioning are related to the comparative intensity of the UCS used during adaptation trials. Should this be the case, conditioning level would be expected to be related to the strength of the adaptation stimulus, number of adaptation trials and intensity of the UCS during conditioning held constant. The present study is concerned with making such a comparison. The condi-

tioning of four groups was studied, a control group given no adaptation trials and three experimental groups given adaptation trials with a noxious stimulus weaker, equal, and stronger in intensity, respectively, to the UCS administered to all *Ss* during conditioning.

Method

Subjects.—The 80 *Ss*, students in introductory psychology courses, were assigned at random to the control or one of the three experimental groups. Each of the four groups consisted of 20 *Ss*.

An additional 12 *Ss* were run in the experiment but were discarded because they failed to meet a latency criterion established in previous studies (**8**) in an attempt to exclude voluntary responders in the conditioning period. Briefly, all records in which more than 50% of the CR's had latencies of less than 300 msec. were excluded from further consideration. Of the *Ss* thus eliminated no relationship was found with group membership.

Apparatus and method of recording.—The CS consisted of an increase in the brightness of a 6-cm. circular milk-glass disc from 1.2 apparent ft.-c. to 23.8 apparent ft.-c. The duration of the CS on each trial was 520 msec. The UCS, a puff of air delivered to the corner of the right eye, was produced by the fall of a column of mercury in a manometer, intensity being varied by length of fall. On conditioning trials the onset of the UCS followed that of the CS by 470 msec.

Before each trial, *S* was given a verbal ready signal. He was instructed to blink once upon receiving this signal and then to fixate on the disc. In both adaptation and conditioning trials, intertrial intervals of 15, 20, and 25 sec. were employed according to a prearranged schedule.

The equipment for presenting the stimuli and recording the eyelid closure is described in detail elsewhere (**7, 8**).

Procedure for adaptation groups.—Preliminary instructions informed *S* that the study was concerned with visual reactions to two types of stimuli administered singly and in combination. The *S* was then given a single presentation of the CS alone, followed by a single presentation of the UCS (30-mm. fall).

The three adaptation groups were then given 50 presentations of the UCS alone. Treatment of the groups differed only in intensity of the UCS employed, the air puff being produced by a 15-mm., 30-mm., and 80-mm. fall of mercury, respectively, for the different groups. Following the adaptation trials, 50 conditioning trials were given, proceeding without interruption or additional instructions. The UCS for all groups was produced by a 30-mm. fall.

Procedure for control group.—Following initial instructions and administration of a single CS and UCS, *S* was given 50 conditioning trials. The procedure and strength of the UCS for these *Ss* were the same as for the experimental groups.

Results

Amplitude in adaptation period.—In order to determine whether the amplitude of the UCR changed over the course of the 50 adaptation trials, the amplitude of each of these responses was measured for the three experimental groups. As Table 1 indicates, the mean amplitude decreased quite consistently over blocks

TABLE 1

Mean Amplitude of Response for Successive Blocks of 10 Adaptation Trials for the Adaptation Groups

Group	Trial Block									
	1		2		3		4		5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
II (15-mm.)	6.97	1.29	6.43	1.16	6.25	1.28	6.11	1.15	5.88	1.02
III (30-mm.)	7.17	1.30	6.85	1.35	6.57	1.29	6.37	1.42	6.05	1.62
IV (80-mm.)	7.35	1.14	7.06	1.14	6.98	1.00	6.77	.99	6.83	1.19

TABLE 2

ANALYSIS OF VARIANCE FOR TREND OF THE AMPLITUDE DATA FOR ADAPTATION TRIALS

Source	*df*	Mean Square	*F*
Between groups	2	11.36	1.71
Between Ss same group	57	6.64	
Between trial blocks	4	6.15	24.90*
Trial blocks × Groups	8	1.02	4.13*
Pooled Ss × Trials	228	.25	

* $P < .01$.

of 10 trials for all groups. Further, in every block of trials the mean amplitude was positively related to the intensity of the UCS. The significance of these results was tested by means of an analysis of variance for repeated measures (**2**), summarized in Table 2. As indicated by the between-trials *F*, response amplitude decreased significantly as the adaptation trials progressed. Comparing the performance of the three adaptation groups, the between-groups *F* fell short of statistical significance. However, the *F* for interaction between trial blocks and groups was significant at the .01 level, the difference in response amplitude between groups which was minimal on the early trials increasing in later trials, the milder puff strengths resulting in more adaptation.

Conditioning.—As may be seen in Table 3, the mean number of CR's was greatest for the control Ss, and for the three experimental groups decreased in magnitude as intensity of the adaptation stimulus increased. An analysis of variance performed to determine the significance of the differences among total number of CR's for the groups yielded a between-groups *F* (4.15) that is significant at the .01 level. In addition, a *t* was computed between the means of the two extreme adaptation groups (15-mm. and 80-mm.) and found to be

TABLE 3

MEAN NUMBER OF CR's FOR 50 CONDITIONING TRIALS

Group	Mean	σM
I (Control)	22.22	3.28
II (15-mm.)	16.36	2.57
III (30-mm.)	12.20	2.19
IV (80-mm.)	9.14	2.28

significant at the .05 level. Thus it may be concluded that intensity of the adaptation stimulus is related to subsequent conditioning performance.

DISCUSSION

Earlier it was suggested that one of the functions of the UCS in defense conditioning is to evoke an internal emotional response, the magnitude of which in part determines drive level. The conditioning results of the present study are consistent with the further hypothesis that the strength of the emotional response and hence of drive to the UCS during conditioning are inversely related to strength of the UCS during adaptation trials.

Several interpretations of the assumed relationship between the hypothetical emotional reaction during conditioning and the intensity of the adaptation stimulus could be offered. First, it is possible that the decreased amplitude of the eye blink of the experimental groups during the adaptation period reflects physiological adaptation of the sensory system affected by the puff of air, the level of adaptation reached being positively related to the magnitude of stimulation. As such sensory adaptation takes place, the magnitude of the emotional response evoked by the [illegible]nulus would also be expected to decrease. While no studies have been reported which have employed the conditions of the present experiment, there is evidence (**1**) that absolute level of adaptation can be a positive function of the intensity of stimulation. Assuming the existence of such a relationship, the conditioning results of the present experiment would be

expected since the magnitude of the emotional response during conditioning, being associated (inversely) with the absolute level of sensory adaptation reached during the adaptation trials, would be least for the 80-mm. group, most for the 15-mm. *S*s, etc.

An alternate hypothesis is that a more general emotional or fear response, such as is assumed to be present in experiments involving mere threat of shock or of some other noxious stimulation, becomes reduced in magnitude during the adaptation trials. That is, *S*s become less fearful of the UCS even though little or no sensory adaptation has taken place. The differences in conditioning performance among the adaptation groups could then be interpreted as being dependent on the relative strengths of the adaptation and conditioning stimulus. Thus the group administered a 15-mm. puff during adaptation experience increased stimulation during conditioning and react with a more vigorous emotional response than *S*s continuing with a UCS of the same strength. Similarly the group changed from 80 to 30 mm. experience a lesser stimulus during conditioning and hence react with an even lower emotional response than at the end of the adaptation period.

Even if the above analysis were correct, some sensory adaptation could be taking place. The more crucial point concerns the results of conditioning: whether performance differences among groups reflect the level of sensory adaptation reached during adaptation trials or whether the relative strength of the adaptation and conditioning stimulus is the critical factor. The present study provides no way to test the two alternatives, although the fact that the amplitude of the response to the air puff is still positively related to the intensity of the puff even at the end of the adaptation trials tends to support the general emotionality hypothesis.

Summary

Four groups were given 50 trials in a defense conditioning situation under the same conditions. Prior to conditioning, three of the groups were given 50 adaptation trials with the UCS alone, the intensity of the noxious stimulus for the groups being weaker, equal to, and stronger, respectively, than the UCS employed during conditioning. The fourth, control, group was given only conditioning trials.

Results indicated that the amplitude of response to the UCS decreased significantly over the course of the adaptation trials for all groups. No significant differences in amplitude were found among groups.

In conditioning, the control group gave the greatest number of CR's, while for the adaptation groups, performance was inversely related to the intensity of the UCS employed during adaptation. These results were interpreted as being consistent with the hypothesis that one function of the UCS during conditioning is to evoke an internal emotional response and, further, that the magnitude of this response is inversely related to the intensity of the stimulus given on the adaptation trials.

References

1. Baker, H. D. The course of foveal light adaptation measured by the threshold intensity increment. *J. Opt. Soc. Amer.*, 1949, **39**, 172–179.
2. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
3. Kellogg, W. N. Electric shock as a motivating stimulus in conditioning experiments. *J. gen. Psychol.*, 1941, **25**, 85–96.
4. MacDonald, A. The effect of adaptation to the unconditioned stimulus upon the formation of conditioned avoidance responses. *J. exp. Psychol.*, 1946, **36**, 1–12.
5. Passey, G. E. The influence of the intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
6. Seward, J. P., & Seward, G. H. The effect of repetition on reactions to electric shock: with special reference to the menstrual cycle. *Arch. Psychol.*, 1934, **27**, No. 168.
7. Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
8. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received March 3, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

THE CORRELATES OF MANIFEST ANXIETY IN STYLUS MAZE LEARNING

HOWARD S. AXELROD, EMORY L. COWEN,[1] AND FRED HEILIZER

University of Rochester

Considerable research interest in recent years has centered around the attempt to extend Hullian learning theory to the study of human motivation. In a substantial number of studies in this area, scores on the Taylor Manifest Anxiety Scale (MAS) have been used as an operational measure of effective drive level, and relations between performance on this scale and performance in varying types of learning situations have been reported.

The MAS, as described by Taylor (**13**), consists of a series of 50 MMPI items selected by clinical psychologists as indicative of manifest anxiety. These items are embedded in a larger scale called the Biographical Inventory. Farber and Spence (**3**), as a rationale for using this instrument in studies of learning, state the assumption that the MAS scores ". . . are an index of [the] *S*s' state of reactivity or excitability which in turn reflects their general level of drive (*D*)." Further, according to Hullian theory, effective drive level interacts multiplicatively with all habit tendencies to determine effective reaction potential. On the basis of this position, the following general predictions about learning ability have been made: (*a*) better performance may be expected from high-anxious *S*s (high MAS score) on a task with a single dominant response tendency, e.g., conditioning, and (*b*) increasing impairment of performance may be expected for high-anxious *S*s on complex tasks, involving competing response tendencies, as a function of the number and strength of competing response tendencies.

Considerable data have now been collected bearing directly on these predictions. In simple conditioning studies, high-anxious *S*s have been found to condition better than low-anxious *S*s (**9, 10, 11, 12**). On complex tasks, the high-anxious *S*s have, in general, been found to perform less effectively (**3, 6, 7, 8, 14**). The foregoing investigations have, for the most part, been based on comparison of extreme high- and low-scale scorers. Some evidence has been presented suggesting that this relationship may not extend throughout the entire A-scale range (**4, 7**).

Problem

The present study is the first of a series designed to examine further the relationship between MAS scores and performance on various tasks. It was designed primarily to replicate the Farber and Spence study on anxiety and stylus maze learning (**3**). These investigators used a 10-choice-point linear T maze with anxious and non-anxious *S*s selected from extreme scorers on the MAS. Each experimental group consisted of 40 students selected from the upper and lower 20% of the distribution, as determined by

[1] This study was supported in part by a research grant to the second author, from the National Institute of Mental Health of the National Institutes of Health of the USPHS (Contract No. MH 437C). This is the sixth in a series of papers emanating from this work. Part of the material reported in the present paper was presented at the Eastern Psychological Association meetings in New York, April, 1954.

a standardization group. It was found that: (*a*) anxious *Ss* were significantly inferior in maze performance, both for trials and errors, and (*b*) anxious *Ss* tended to make relatively more errors at the more difficult choice points.

The present study introduces two methodological extensions not present in the Farber and Spence study: (*a*) As a further examination of the sensitivity of the MAS over the whole range of scores, a middle-anxious group (43rd to 56th percentiles) was included in the present experiment. (*b*) In the Iowa studies, *Ss* with Lie-Scale scores of seven and above were eliminated from experimental consideration.[3] It was felt that high Lie scorers may not have been responding to the MAS items objectively. This added selection criterion was not used in the present study, since our intention was to utilize Lie-Scale data as an integral part of the statistical analysis. In fact, since high Lie scorers tended to fall primarily in the low MAS range, an attempt was made to select a relatively higher proportion of low-anxious *Ss* for purposes of the Lie-Scale analysis.

Method

Subjects.—Ninety-six freshman men and women were chosen from a group of 276 on the basis of performance on the MAS. Three groups were selected: High anxiety—scores of 27 (90th percentile) and above; Middle anxiety—scores of 12 to 14 (43rd to 56th percentiles); Low anxiety—scores of 6 (20th percentile) and below. Subjects in each group were: 27 high (14 male and 13 female), 30 middle (17 male and 13 female), and 39 low (26 male and 13 female). Subjects were seen in individual testing sessions on the basis of random assignment to one of the three *Es*.

Apparatus.—The stylus maze was of the same design as that described by Farber and Spence. The only variations from their model were in the use of sheet aluminum instead of brass and a very slight alteration of the over-all dimensions of the plate into which the maze pattern was cut, which in no way affected the maze-pattern dimensions.[3]

Procedure.—The present procedure was essentially the same as in the study being replicated. Two slight changes in the present study involved: (*a*) the use of an *automatically timed*, 5-sec. interval, signal (buzzer) for moving to succeeding choice points, and (*b*) the use of opaque goggles instead of a curtain screen to prevent visual examination of the maze.

Prior to the start of the experiment, all *Ss* were given several practice trials on a three-choice-point demonstration maze, similar in general design to the experimental maze, in order to familiarize them with the use of the buzzer signal and the verbal instructions at the choice points.

The experimental procedure consisted of having *Ss* move from choice point to choice point at 5-sec. intervals indicated by the buzzer signal. When *S* made an incorrect choice, thereby moving toward a blind alley, he was told "no." He then reversed direction until he was just past the choice point (now pointing in the correct direction) when he was told "stop." If a correct choice was made, *S* was told "yes," and he stopped immediately. In either case, after he stopped, *S* waited for the buzzer and then proceeded to the next choice point where the directions "no" and "stop," or "yes," were again given. This procedure was repeated for all choice points. The criterion of mastery was two consecutive errorless trials.

Results

MAS analysis.—Means and *SD*'s for all MAS groups for error scores and trials to criterion are presented in Table 1. The analysis of variance[4] for error scores and trial scores

[3] Personal communication from Dr. I. E. Farber.

[3] According to a personal communication from Dr. I. E. Farber, the correct choice at the last choice point is to the right and not, as reported in the Farber and Spence paper, to the left.

[4] Because there were unequal cell entries, the analysis of variance used is an approximation method (5) requiring that the following two assumptions be satisfied: (*a*) homogeneity of variance, and (*b*) random distribution of cell entries. For all A-scale analyses homogeneity of variance was tested by means of Bartlett's test, and the random distribution of cell entries was tested by a chi-square technique. In each case it was possible to demonstrate that the prerequisite assumptions had been fulfilled.

TABLE 1

MEANS AND *SD*'S FOR STYLUS MAZE PERFORMANCE OF HIGH, MIDDLE, AND LOW ANXIETY GROUPS

Group	*N*	Errors		Trials	
		Mean	*SD*	Mean	*SD*
High	27	34.8	16.3	13.6	7.5
Middle	30	30.4	18.8	12.4	7.9
Low	39	29.3	14.4	12.7	7.5

is presented in Table 2. In the case of error scores, the Anxiety-Sex-Experimenter interaction proves to be highly significant when tested over the within-error term. In order to use the significant triple interaction as an error term, at least one of the variables must be randomly distributed (**2**). Neither the Anxiety nor the Sex variable meets this criterion and, since we are unwilling to assume *post hoc* a random selection of *E*s, the within-error term was used to test the double interactions and the main effects. Thus, in Table 2, it may be seen that, although none of the main effects reach the .05 level of confidence, all the interactions are significant at the .05 level of confidence or beyond when tested over the within-error term. In order to increase the clarity of the results, separate analyses were done for each sex. Neither the Anxiety main effect nor the Anxiety-Experimenter interaction reaches statistical significance for either sex. The Experimenter main effect, however, is statistically significant for men, although not for women.

TABLE 2

ANALYSIS OF VARIANCE FOR ERROR AND TRIAL SCORES

Source	*df*	Error Scores		Trial Scores	
		MS	*F*	*MS*	*F*
Anx.	2	499.8	1.74	166.8	2.76
Exp.	2	182.2	—	68.3	1.13
Sex	1	887.7	3.09	153.5	2.54
A×E	4	797.6	2.71*	147.7	2.45
A×S	2	1180.0	4.10*	170.5	2.82
E×S	2	2778.2	9.66***	496.4	8.22***
A×E×S	4	1410.8	4.91**	33.3	
Within	78	287.6		60.4	
Total	95				

* $P = .05$.
** $P = .01$.
*** $P = .001$.

In the case of the trial scores, only the Experimenter-Sex interaction reaches statistical significance when tested over the within-error term. Once again separate analyses of variance were done for each sex. Only the Anxiety-Experimenter interaction in the analysis of variance for men reaches the .05 level of confidence. (In the case of both the error and trials-to-criterion analyses, since the analysis of variance used is an approximation method, the precision of which is not known, *t*'s were computed between high, low, and middle MAS groups in all combinations. None of the *t*'s approached significance at the .05 level of confidence. In order to make our results entirely comparable to the Farber and Spence data, these *t*'s were also computed eliminating all L-scale scores of 7 and above. The results were not altered by this analysis.)

Choice-point analysis.—In the Farber and Spence study the prediction was made that high MAS scorers would perform progressively worse on the more difficult choice points as compared to low MAS scorers. The level of difficulty for each choice point was determined by the number of errors made by an independent, unselected group of students from undergraduate experimental psychology classes. A rank-order correlation was computed between choice-point difficulty level and the total errors made by high MAS scorers minus the total errors made by low MAS scorers

TABLE 3

DIFFICULTY LEVEL OF CHOICE POINTS AND MEAN ERRORS AT EACH CHOICE POINT FOR HIGH AND LOW ANXIETY GROUPS

Group	*N*	1	2	3	4	5	6	7	8	9	10	Total
Difficulty*		5	1	8	3	10	4	9	2	7	6	
High	27	2.78	1.93	5.52	1.59	8.04	2.52	6.37	1.04	2.81	2.22	34.81
Low	39	2.26	1.90	3.51	1.15	6.51	2.03	5.74	1.36	2.41	2.44	29.31

* Higher ranks represent greater difficulty.

at each choice point. The resultant rho of .74 was interpreted to mean that the high MAS scorers performed progressively worse, as compared to low MAS scorers, as a function of the level of difficulty of the choice points. The above procedure and analysis was reproduced in the present experiment. The data presented in Table 3 contain the choice-point difficulty level, as determined by a new standardization group of undergraduate experimental psychology students at the University of Rochester ($N = 24$), and the mean number of errors made at each choice point by high and low MAS scorers.

While the resulting rho of .76 is in direct agreement with the comparable Farber and Spence data, there is a serious objection to the manner in which this type of rank correlation is computed. In the Farber and Spence experiment, the high MAS group made significantly more errors for the maze as a whole than the low MAS group. In the present experiment, the high MAS group also made more errors than the low MAS group, although this difference did not reach statistical significance. It is to be expected, therefore, that the positive correlations reported above were, at least in part, an artifact of the greater number of errors made by the high MAS group over all the choice points. Thus, when absolute differences in errors at the various choice points are used, the discrepancy between groups is spuriously accentuated on the more difficult choice points compared to the easier ones.

In order to control for the total

TABLE 4

PERCENTAGE OF TOTAL ERRORS AT EACH CHOICE POINT FOR HIGH AND LOW ANXIETY GROUPS

Group	Choice Points										
	N	1	2	3	4	5	6	7	8	9	10
Farber & Spence											
High	40	4.2	5.1	8.6	8.6	16.2	6.5	17.4	6.5	10.8	16.1
Low	40	4.0	4.7	7.3	7.3	16.1	9.7	18.2	5.7	10.6	16.4
Axelrod, Cowen, & Heilizer											
High	27	8.0	5.5	15.9	4.6	23.1	7.2	18.3	3.0	8.1	6.4
Low	39	7.7	6.5	12.0	3.9	22.2	6.9	19.6	4.6	8.2	8.3

number of errors made by each group, rank-order correlations were computed, for both experiments, comparing the difficulty level of the choice point with differences between the two groups in percentage errors at each choice point (see Table 4). When this was done, the obtained rhos for the Farber and Spence study and for this experiment, respectively, were $-.50^{6}$ and $+.32$ which for 9 *df*, do not reach the .05 level of confidence.

The above analysis is one method of controlling for differences in performance between groups (i.e., total number of maze errors) in this particular experimental situation. Farber and Spence, however, attempted also to control for this factor by selecting 20 pairs of anxious and nonanxious *Ss* with the same total error score and comparing the errors for the two groups on the five easiest and five most difficult choice points. The resulting *t* of 1.84 was significant at the .05 level of confidence (one-tailed test). A similar analysis was done in the present study using 18 pairs of *Ss*. In this case 9 pairs had identical scores and 9 pairs differed by one error. The data for these groups of paired *Ss* are presented in Table 5. The *t* for matched pairs yielded a value of .29 which, for 17 *df*, does not approach significance at the .05 level of confidence.

TABLE 5

ERROR SCORES OF 18 PAIRS OF HIGH AND LOW MAS SCORERS (MATCHED FOR TOTAL ERRORS) AT THE FIVE LEAST DIFFICULT AND THE FIVE MOST DIFFICULT CHOICE POINTS

Group	Least Difficult		Most Difficult	
	Mean	*SD*	Mean	*SD*
High	7.89	3.91	21.17	9.68
Low	8.00	3.55	20.83	9.76

[6] This figure is based on Farber and Spence's reported difficulty level. Using choice-point difficulty levels as established with the present standardization group, the rho becomes $-.35$.

TABLE 6

STYLUS MAZE PERFORMANCE FOR HIGH AND LOW LIE GROUPS, FOR THE TOTAL POPULATION AND WITHIN THE LOW ANXIETY GROUP

Group	*N*	Errors		Trials	
		Mean	*SD*	Mean	*SD*
Total High Lie	25	36.6	18.49	15.9	9.59
Total Low Lie	42	26.6	15.50	10.5	6.02
High Lie-Low A	15	31.7	13.75	14.9	8.78
Low Lie-Low A	14	24.6	12.93	9.4	5.54

Lie-Scale analysis.—It was also our intention to test for differences between high and low Lie groups for trials to criterion and total errors by means of the above analysis-of-variance technique. Unfortunately, however, one of the basic conditions of this technique—random distribution of cell entries—was rejected at the .05 level by means of a chi-square test. The *t*'s were computed for trials and for errors between high (scores of 6 and above) and low (scores of 3 and below) Lie groups for the total experimental population, and also between high and low Lie scorers within the low MAS group. The data for these analyses are presented in Table 6. The *t* ratios between the total sample high and low Lie groups for Trials and Errors were significant at the .05 level of confidence or better, using a two-tailed test ($t = 2.79$ and 2.34, respectively). Although the differences between high and low Lie groups within the low MAS group were in the same direction and of approximately the same magnitudes for both trials and errors, as the differences between the total population high and low Lie groups, the resulting *t*'s of 1.93 and 1.38, respec-

tively, did not reach the .05 level of significance due to the reduced *N*'s.

Choice-point analyses (structurally similar to those done for anxiety groups) were also computed for high and low Lie groups for the total sample and within the low MAS group. The resulting rhos of +.39 for the total population and +.42 for the low MAS group do not reach the .05 level of significance.

Discussion

MAS analysis.—No significant differences among MAS groups were found on total errors or trials to criterion. It is interesting to note that the interactions seem to be more important in these analyses than are any of the primary effects, and that the Sex and Experimenter variables seem to be at least as important as the Anxiety variable. This is particularly noteworthy because previous studies in this area have tended to ignore the former variables and because there is, as yet, no place for them in the deductions tested here.

A finding of special interest is the difficulty level of the stylus maze per se in the present experiment. Here *S*s required almost twice as many trials and made nearly twice as many errors as did *S*s in the Farber and Spence study. This finding is surprising since: (*a*) in both studies the *S*s were college students, and (*b*) the distribution of MAS scores, upon which the selection of *S*s in the present study was based, was almost identical to the population distribution from which *S*s were selected by Farber and Spence. Thus it is not likely that differences between the two groups in intelligence or in measured anxiety level would be large enough to account for performance differences in the two studies. In trying to account for these differences, two procedural variations warrant consideration. The first is our use of opaque goggles to prevent seeing the maze in comparison to the curtain screen that Farber and Spence used. Probably opaque goggles are more effective in reducing visual and tactual cues than is a curtain screen, thus making for a higher difficulty level in maze performance in the present experiment. It is also possible that the more complete reduction of normal, everyday cues tends to disorganize naive *S*s, thus impairing their manifest performance.

A second consideration is the nature of the samples used in the two studies. Farber and Spence utilized *S*s from psychology classes, while in the present study, *S*s were recruited from freshman English classes and had not taken any psychology courses. Our *S*s may, therefore, have been more threatened by the experimental situation than were the less naive *S*s utilized by Farber and Spence, this in turn resulting in less efficient performance.

Whether these two hypotheses can plausibly account for the striking differences in performance level between *S*s in the two experiments is a matter of personal choice until tested experimentally. The deductions from Hullian theory would, if anything, lead us to expect even more marked differences between MAS groups under present experimental conditions than were reported by Farber and Spence. Actually the opposite results were obtained.

One final consideration remains. Despite our essentially negative findings with respect to anxiety level and maze performance, there appears to be at least a slight tendency for the high-anxious group (only when compared to the low-anxious group) to require a greater number of trials and to make more errors on the maze. Thus, the direction of our findings in this particular phase of the study is similar to that reported by Farber and Spence at a nonchance level of confidence. We can at best only speculate as to the interpretive significance of this state of affairs. There is, of course, the distinct possibility that the present findings represent nothing more than chance departures, and that the similarity in the direction of results in the two experiments therefore carries no special interpretive significance. If this is not so, some resolution for the disparate findings may possibly reside in two areas. First, the relations between drive level and performance may be more complex than present formulations suggest, and second, the MAS may be an imprecise measure of drive level (i.e., its measurement properties may change with changes in administrative conditions, motivation of *S*s, etc.).

Choice-point analysis.—When performance was held constant, no significant relationships were found between choice-point difficulty and differential perform-

ance on the part of MAS groups. This result holds true for both methods of controlling for performance differences used in the present study: (*a*) computing the rank-order correlation on the basis of percentage of error scores at each choice point, and (*b*) pairing high and low MAS *Ss* on the basis of total errors and comparing the errors for the two groups on the five easiest choice points, by means of a *t* test for matched pairs. The writers believe that the first method is preferable to the second method, because it uses the entire sample while the latter uses only a part of the total sample and may introduce selective sampling. For example, when the first method was applied to the Farber and Spence data, a rank-order correlation of −.50 was obtained, indicating that the low MAS group performed progressively worse, as compared to the high MAS group, as the difficulty level of the choice points increased. This correlation, opposite in direction to the predicted relationship, describes the entire group of 40 high MAS and 40 low MAS *Ss*. On the other hand, when Farber and Spence selected 20 pairs of high and low MAS *Ss* with each pair matched as to total error score, and compared their relative performance on the five easiest choice points, the resulting *t* describes only one-half of the entire group of *Ss*. For these results, since the statistic describing the entire group indicates a negative relationship, while the statistic describing a part of the total group indicates a positive relationship, one may raise the question of whether the latter result derives from a nonrepresentative sample of high- and low-anxious *Ss*.

This question increases in importance upon further examination of the Farber and Spence data. The mean error scores for the high and low MAS groups, respectively, were 19.15 and 13.70, while the mean error score for the 20 pairs of matched *Ss* was 13.55. These data seem to indicate that the high MAS subgroup selected by Farber and Spence represents an atypical sample.

Thus it would seem that neither the Farber and Spence choice-point analysis nor the data of the present study lend support to the pertinent deductions from Hullian theory.

Lie-Scale analysis.—Significant differences were obtained between high and low Lie-Scale scorers for the whole sample on total errors and on trials to criterion, indicating that the low Lie group performed better than did the high Lie group. These results might be considered to support the Farber and Spence position only if the assumption is made that the Lie Scale measures drive level. Lacking further evidence on this point, however, and because the Lie Scale was not devised as a measure of drive level, we may interpret this result either as a chance occurrence or as one not yet integrated within any particular theoretical system.

Choice-point analyses were computed for high and low Lie groups, both for the total population and within the low MAS group. It was felt that if these results were also consistent with the deductions from Hullian theory (substituting Lie for Anxiety: high Lie scorers would perform progressively worse on the more difficult choice points as compared to low Lie scorers) then it would be more justifiable (*a*) to argue that the Lie Scale might represent a type of measure of drive level, and (*b*) to interpret the Lie-Scale analyses as supporting the Farber and Spence theoretical position. The resulting rhos, however, did not reach the .05 level of confidence.

Summary

The present study replicated the Farber and Spence investigation of the relationship between scores on the Taylor Manifest Anxiety Scale and stylus maze performance. On the basis of deductions from Hullian theory, it had been predicted that: (*a*) high MAS scorers would perform more poorly on the maze than would low MAS scorers, and (*b*) high MAS scorers would perform progressively worse on the more difficult choice points, as compared to low MAS scorers. Additionally, in the present study, a group of *Ss* whose MAS scores fell in the middle range was included. Subjects were selected in such a way as to make possible a regrouping and secondary analysis on the basis of Lie-Scale data.

Ninety-six male and female *Ss* were selected from a group of 276 freshmen, as high, middle, and low MAS scorers, and were tested on the stylus maze. In contrast to the Farber and Spence findings, no significant differences were found among anxiety groups with respect either to total errors or trials to criterion. No significant relationships were found between choice-point difficulty and differential performance between high and low anxiety groups. A re-examination of the comparable Farber and Spence data suggests that their conclusions may be based on an inappropriate method of analysis. With respect both to over-all maze performance and choice-point analyses, our data do not support either the Farber and Spence findings or present simple deductions from Hullian theory.

Regrouping our data, it was found that low scorers on the Lie Scale perform significantly better than high Lie scorers both with respect to trials to criterion and total errors. Differences in maze performance between "Lie" groups are greater than differences between anxiety groups. These empirical findings are suggestive rather than conclusive, and warrant further research consideration.

References

1. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
4. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
5. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
6. Lucas, J. D. The interactive effects of anxiety, failure and intraserial duplication. *Amer. J. Psychol.*, 1952, **56**, 59–66.
7. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
8. Ramond, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
9. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
10. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.
11. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
12. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
13. Taylor, J. A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
14. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **45**, 61–64.

(Received February 17, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

DEVELOPMENT AND MAINTENANCE OF THE PREFERENCE VALUE OF AN OBJECT [1]

HAROLD M. SCHRODER

The Ohio State University[2]

The development of goal-directed behavior and particularly the maintenance of goals under varying conditions continue to constitute important problems for personality theory. This study is concerned with the measurement and relationship between two variables, having relevance to the development and maintenance of the preference or reinforcement value of an object.

It is well known that a neutral object takes on reinforcing qualities when it is paired with a primary reward (**1, 7, 8, 11**). In all these experiments the strength of the learned reward was determined by maintenance measures: for example, trials to extinction or the learning of a new problem on the basis of the learned reward. Some studies employing maintenance measures have been concerned with variation in the strength of a learned reward as a function of the reinforcement ratio during training (**3, 6, 10**). In contrast to these investigations another group of studies has utilized the measure of preference or reinforcement value of an object (**2, 4, 5**). The preference value of an object is measured by a choice technique relative to other objects having known preference values.

Although no explicit rationale about the relationship between these two measures, namely maintenance and choice measures, is presented, it is assumed that they are independent, that is to say, the preference value of an object is not expected to have any necessary relationship to the maintenance of that value under conditions of extinction. The present study aims to examine this relationship and has two aspects. First, it aims to investigate the relationship between the reinforcement ratio during training, and preference value measured immediately after training. It is hypothesized that the preference value of an object is a function of the probability that it will be followed by a subsequent reinforcment and the value of the subsequent reinforcement (**9**). Partial training would be expected to result in a lower preference value. Second, an initial step will be taken to explore the relationship between preference value and the maintenance of that value during extinction. It is predicted that, although objects will have a lower preference value after partial training, the value of these objects will be maintained over a greater number of extinction trials.

METHOD

Subjects.—The *Ss* were 60 female children from the second, third, and fourth grades.

Materials.—A manually operated vending machine was constructed in which *S* could insert tokens and then pull a lever so that a charm was ejected. The tokens consisted of ¾-in. square cardboard pieces, and 20 different charms were used as rewards.

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology at The Ohio State University. The author is greatly indebted to Professors J. B. Rotter and D. D. Wickens for their valuable aid during various phases of this study.

[2] Now at Princeton University.

Procedure.—Each *S* was presented with 20 different charms and one token and asked to select the most preferred charm. This procedure was continued until the charms were in order of preference from highest to lowest. The charms that ranked 1, 2, 4, 5, 7, and 8 were selected as reward charms. They were used as subsequent rewards to pair with the token. The charms that ranked 3, 6, 9, 12, 15, 18, and 21 were selected as the scale charms, since they represent a six-step preference scale for each *S*. In all cases the token was ranked 21. For these steps to represent meaningful differences in reinforcement value, the reliability of *S*s' rankings must be high. Therefore, *S* was asked to rerank the six scale charms and the token, allowing the computation of reliability.

Training.—The instructions were as follows: "You have always placed this [pointing to the token] at the bottom, but I will show you how it can win prizes for you. Every time you get one of these [tokens] you may place it in here [pointing to slot in vending machine] and pull this lever, and if anything comes out here [pointing to ejection slot] you may keep it." The *S* was handed three tokens at a time and the instructions were repeated on the occurrence of every second reward. The order of the occurrence of charms was the same for all *S*s in terms of the rank position of the charm.

Groups.—Group 1 received 8 rewards over 16 trials. The order of the reinforcement was ++−+−−+−++−−+−+−. Group 2 received 8 trials and 8 rewards, while Group 3 received 16 trials and 16 rewards. That is, for two groups (1 and 2) frequency of reinforcement was held constant and the probability of reinforcement was varied, and for two groups (2 and 3) frequency of reinforcement was varied and the probability of reinforcement was held constant.

Determination of RV of token.—As soon as the training was completed, *E* stated: "Let us stop here a while. I want you to make some choices for me and whatever you choose you may keep. [placing a token beside the lowest ranked charm, No. 18]. If you select this one [pointing to token] you may place it in the machine and if anything comes out you may keep it, and if you choose this one [pointing to charm] you may put it with your prizes and keep it." This procedure was continued up the preference scale until all choices were made. The more times the token was chosen instead of à charm, the higher the RV of the token.

Extinction.—Since there were slight variations in the number of tokens accumulated under different conditions, *E* removed the tokens accepted by *S* to a position where they were not visible to *S* and merely handed *S* three tokens at a time during extinction. After every three extinction trials the RV of the token was determined as described above. Extinction was continued until no charms remained, i.e., until the token returned to its original RV, the lowest in the scale.

Results

The median rho between the first and second ranking of the charms was .95, so that it may be assumed that the six scale charms represented meaningful steps on an empirical scale for the calculation of preference value of the token.

Preference value after training.—The mean increase in preference value of the token for Groups 1, 2, and 3 was from zero to 3.5, 4.40, and 4.75, respectively, before extinction began (see Fig. 1). The 50%-reinforcement group ascribed a significantly lower preference value to the token than both continuous-reinforcement groups. The difference between the groups given 8 and 16 continuous reinforcements was not significant.

Maintenance of preference value.—Figure 1 presents curves for the value of the token after every three extinction trials. Although the token had a significantly lower preference value for Group 1 (50%) after training, Group 1 needed significantly more extinction trials for the token to reduce in value to its original position of

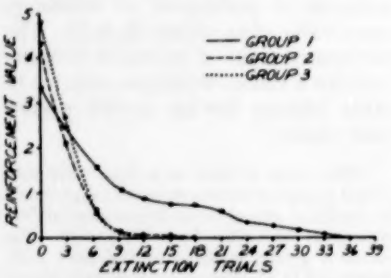

Fig. 1. Reduction of preference value as a function of extinction.

zero on the scale. After the sixth extinction trial the token had a significantly higher preference value for Group 1 than for either of the two continuous groups. The mean number of blocks of three extinction trials taken for the token to return to its original preference value was 5.40, 2.85, and 2.40 for Groups 1, 2, and 3, respectively. The 50% group required significantly more extinction trials than Groups 2 or 3.

Discussion

The results are in essential agreement with the hypotheses. When a choice method is used as a consequent condition for a learned reward, its value would appear to be some function of the probability that it was followed by a subsequent reward during training. When partial and continuous training sequences are compared, the original lower preference value after partial training maintains that value over significantly more extinction trials. Under these training conditions, the strength of a given goal-directed behavior is unrelated to its persistence under extinction conditions. Although most psychologists assume this experimental finding, the implications have not been generally utilized or subjected to further research. The question of maintenance still offers a challenging problem since an adequate description would appear to involve the nature of the thought processes or assumptions developed about sequential events.

Summary

Using 60 elementary school children, a previously neutral token was associated with rewards so that expectancy and frequency varied independently. Group 1 received 16 trials and 8 rewards, Group 2 received 8 trials and 8 rewards, and Group 3 received 16 trials and 16 rewards. The reinforcement value of the token was measured by a choice technique relative to other objects having a known scale value. Measures of reinforcement value were taken after training and after every three extinction trials. The token had a significantly lower reinforcement value for Group 1 (partially reinforced) than for Groups 2 and 3 (continuously reinforced), but maintained that value over significantly more extinction trials. No differences were found between Groups 2 and 3.

References

1. Cowles, J. T. Food tokens as incentives for learning by chimpanzees. *Comp. Psychol. Monogr.*, 1937, **14**, No. 5.
2. Dunlop, R. L. Changes in children's preferences for goal objects as a function of differences in expected social reinforcement. Unpublished doctor's dissertation, Ohio State Univ., 1953.
3. Ehrenfreund, D. The effect of partial reinforcement on secondary reinforcing cues. *Amer. Psychologist*, 1948, **3**, 241 (title only). Reported in W. O. Jenkins and J. C. Stanley, Partial reinforcement; a review and critique. *Psychol. Bull.*, 1950, **49**, 193–234.
4. Filer, R. J. Frustration, satisfaction and other factors affecting the attractiveness of goal objects. *J. abnorm. soc. Psychol.*, 1952, **47**, 203–212.
5. Hunt, D. E. Changes in goal-object preference as a function of expectancy for social reinforcement. *J. abnorm. soc. Psychol.*, 1955, **50**, 372–377.
6. McClelland, D. C., & McGowen, D. R. Effect of variable food reinforcement on the strength of a secondary reward. *J. comp. Psychol.*, 1953, **46**, 80–86.
7. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.
8. Pavlov, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927.
9. Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1955.
10. Saltzman, I. J. Maze learning in the absence of primary reinforcement; a study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.
11. Williams, K. A. The reward value of a conditioned stimulus. *Univ. Calif. Publ. Psychol.*, 1924, **4**, 31–55.

(Received February 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

THE EFFECT OF EGO ORIENTATION AND PROBLEM DIFFICULTY ON MUSCLE ACTION POTENTIALS [1]

MARY E. REUDER

Queens College

The present investigation was undertaken to ascertain the comparative effects of ego- and task-orienting instructions on muscular tension behavior during problem-solving activity. It was hypothesized that a significant difference in amount of muscular tension would be found between ego-instructed and task-instructed *Ss* when both groups were engaged in the same task.

This hypothesis resulted from an attempt to relate studies which deal with the concept of psychological tension with those involving physiological muscular tension. No clearly definable operations have enabled investigators to specify psychological tension as an independent variable. Nevertheless, it has been postulated as a variable which accounts for the facilitative or inhibitive effects of ego-orienting (or esteem-threatening) instructions as contrasted with task-orienting instructions. Similar effects on behavior have been observed when physiological muscular tension has been directly measured and/or systematically varied. Both variables, the one a hypothetical construct and the other a directly measurable dimension of performance, are described as affecting behavior in terms of a continuum with disruption occurring at the extremes and optimal facilitation of performance somewhere in the middle range (compare **2, 3, 4, 5, 15, 16, 19, 20, 21, 22, 23, 25, 31** with **6, 7, 8, 9, 10, 18, 24, 27, 28, 29, 30**).

Traditionally, instructions in muscular tension studies have been task-orienting and have not constituted an independent variable. The most relevant evidence available showing a relationship between psychological tension-arousing conditions and physical activities is Freeman's analysis (**12**) of patellar tendon deformation and Smith's electromyographic study (**29**) with the interrupted-task technique.

METHOD

Design.—A factorial design for two independent variables was employed. The independent variables were: type of instructions (ego-orienting vs. task-orienting) and difficulty of the problems (easy vs. hard). The major dependent variable was muscle action potential (MAP) changes during problem solution. Records were also kept: of MAP changes during administration of instructions and during between-problem rest; of verbal responses of *S* to the problems; and of length of time (up to 1 min.) required by *S* to report a solution. Initial MAP levels obtained during a preliminary rest period served as a control measure. There were thus four sets of experimental conditions: Ego-easy (E-E), Ego-hard (E-H), Task-easy (T-E), and Task-hard (T-H).

Apparatus.—The recording apparatus consisted of a 3-stage resistance-capacitance single-ended amplifier, the response of which is flat from below 25 cps to 1000 cps after which it falls off approximately 3 db per octave; a surface electrode attached to the outer side of *S*'s upper right forearm; a grounded electrode resting on a

[1] This report is based on a dissertation submitted in 1951 to the Graduate School of Arts and Sciences, University of Pennsylvania, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. An abridged version was read before the American Psychological Association, September 1953. The author is indebted to Dr. William A. Shaw for his criticisms and suggestions during the planning and execution of the study; to Dr. Marvin A. Iverson both for his assistance in acting as E_2 and for his independent check of all calculations; and to Mr. George Croswell for his advice and assistance on the apparatus.

felt pad soaked in saline solution attached on the inner side of the elbow of the same arm. A DuMont 5-in. cathode-ray oscilloscope and a 16-mm. Bell and Howell moving picture camera set to a speed of 16 frames per sec. were also part of the recording apparatus. A General Radio microvolter in connection with a General Radio beat frequency oscillator was used to calibrate the recording apparatus at a 10μv level. All of this apparatus was housed in one room adjacent to the shielded metal cage in which *Ss* were placed.

Subjects.—It is known that on problem-solving tasks there is an inverse relationship between MAP changes and intelligence (**24, 28**). For this reason, *Ss* were matched in groups of four on SAT scores and randomly assigned to one of the four sets of experimental conditions. The final data were based on a total of 40 *Ss* selected from white, male, introductory psychology students.

Experimental problems.—Number progressions from the Hunt Arithmetic Progressions Test (**17**) were used as the problem-solving task. The easy-problems task consisted of Hunt's Group VII (e.g., 2—3—5—8—12 ——, ——). The hard-problems task was Hunt's Group IX (e.g., 2—3—5—2—6—11—5 ——, ——). Because of the possible within-group variation in difficulty of individual problems (the test is not standardized), the five problems of a group taken together were considered as a single task in evaluating the results. All of the problems used were mathematically solvable.

Instructions.—General introductory instructions[2] to relax, close his eyes, and make no extraneous movements were presented to each *S*. He was further told that because it was necessary in this experiment for all *Ss* to think about the same thing, arithmetic problems would be presented. Ego-oriented *Ss* were further instructed that the problems comprised a short-form intelligence test which would be used in further evaluation of them. Task-oriented *Ss* were told that the problems were unstandardized and that *E* was merely trying to find out something about the problems. All groups of *Ss* were given the same four practice problems (see **26**) prior to presentation of the experimental tasks. At various points throughout the instructions all *Ss* were cautioned against extraneous movement and requested to relax physically as much as possible.

Experimental procedure.—Two *Es* were necessary to carry out the procedure, one to run the recording equipment and the other to administer the instructions to *S*. As each *S* arrived, E_1 tried to put him at ease by introducing him to E_2, by showing him around the laboratory, and by giving him a brief explanation of the equipment and shielded cage necessary for recording muscular activity.

As soon as *S* appeared to be somewhat at ease, E_1 instructed him to relax in the chair located in the shielded cage. As she attached the electrode, E_1 reiterated that it was impossible for *S* to receive a shock from the apparatus. When E_1 finished the introductory relaxation instructions, E_2 entered the shielded cage and remained with *S* throughout the rest of the procedure. E_1 ran the recording equipment in the adjacent room.

When *S* had reached a resting level of 10μv or below, E_2 gave the appropriate experimental instructions which were followed by the practice problems. He then presented the experimental problems one at a time by placing the problem card on a stand which had been previously adjusted to *S*'s reading eye level. Each response was timed with a stop watch from the moment of exposure of the card to the time *S* closed his eyes as a signal that he had solved the problem. E_2 then requested and recorded *Ss* verbal solution. Because E_2's voice was audible in the recording room, E_1 could coordinate recording procedures with experimental procedures.[3]

Recording procedure.—A standard series of photographic records of the oscilloscope tracings was made for each *S*. This sequence was as follows: (*a*) Initial and final calibration recordings of the 10μv input of the oscillator and microvolter were taken for approximately 1 sec. They were taken at the beginning and end of each *S*'s session to control gain of the amplifier. (*b*) Two initial resting-level records of approximately 1½ sec. each and 30 sec. apart were taken after *S*'s tension had dropped to or below 10v. (*c*) Instructions-level records were taken for 3 sec. following the onset of instructions. Five seconds later a series of 1½-sec. samples were taken every 3 sec. and continued until the sample problems were reached in the instructions. (*d*) Work-level records were timed slightly differently for the easy- and hard-problem *Ss*. For hard-problem *Ss*, 1½- to 2-sec. samples were taken every 5 sec. from exposure of the problem and continued either until a solution was given or 60 sec. had elapsed. Since easy-problem *Ss* could often solve the problems in less than 5 sec., the first 3 to 5 sec. of their work time was con-

[2] The exact wording of all instructions is presented in full in (**26**).

[3] At the end of the session, each *S* was informed of the true nature of the experiment and asked not to discuss it with others until after a given date. Care was taken to assure the ego-oriented *Ss* that the tasks were part of an unstandardized test and could tell little concerning their ability.

tinuously recorded. After this, 1½- to 2-sec. samples were taken in the same way as for hard-problem *Ss*. (*c*) Between-problem rest-level records were taken with the same duration and spacing as those for initial rest. Immediately after *S* had given his solution to a problem he was requested to relax. Recordings were made as soon as the MAP level reached approximately 10μv. For *Ss* whose MAP did not drop this low, records were begun at the end of 3 min. of rest.

Measurement.—MAP records were projected against a cardboard scale marked off in centimeter distances. The projector was adjusted so that the peak-to-peak distances of the 10μv calibration records covered 10 cm. on the screen. The widest peak-to-peak distance in each frame of film was then measured in centimeters and directly recorded in microvolts. Average work measures were obtained for *S* by totaling all frames of his five work periods and dividing by the total number of measures made. A similar procedure was employed for the initial rest levels, instruction period levels, and the between-problem rest levels.

Average time per problem was calculated for each *S* by totaling his working time and dividing by five. Accuracy scores were simply the total number of completely correct solutions.

Results

Muscle potentials.—Means and *SD*'s of all the MAP measures are presented in Table 1. The experimental design required that the four groups of *Ss* be homogeneous with respect to the initial, criterion, rest level. Bartlett's test for homogeneity of variance gave a chi square of 3.67 uncorrected for size of sample. Since the correction would have increased further the obtained P value of $> .20$, the null hypothesis was accepted and these rest measures were assumed to be homogeneous. An analysis of variance of these data produced no F ratios significant at the .05 level or above. The four experimental groups of *Ss* were, therefore, considered equivalent in meeting the initial rest-level criterion.

MAP's taken during instructions, between-problem rest, and during work all showed an increase over initial rest measures. These increases in response under experimental conditions for individual *Ss* were correlated with individual initial rest measures. Moreover, the degree of correlation varied from one set of experimental conditions to another despite the established homogeneity of the initial rest-level criterion. For each of the three sets of MAP measurements, the chi square for homogeneity of r showed heterogeneity significant beyond the .02 level.

To rid the various experimental measures of the correlation with initial rest, the r for each cell condition with criterion rest was determined. Then, for each raw score, the appropriate b weights were used to obtain predicted scores. Finally, by subtracting the predicted from the obtained scores, final adjusted scores were computed for each *S*. These

TABLE 1

Means and *SD*'s of MAP Changes in Microvolts
(N = 10 per group)

Experimental Group	Initial Rest (Criterion)		During Instructions		Between-Problem Rest		Work on Problems	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
E-E	5.27	5.89	6.25	7.19	6.66	7.96	8.01	10.08
E-H	4.07	4.56	4.61	5.10	4.18	4.62	5.08	5.58
T-E	4.22	4.56	4.86	5.34	4.58	4.98	5.37	5.87
T-H	4.84	5.29	8.27	12.22	4.84	5.32	7.62	9.04

scores represented the effects of the experimental conditions with individual differences in initial resting level removed. Analysis of variance of these adjusted scores for instructions and for between-problem rest yielded no significant *F* ratios. Any apparent differences in the unadjusted, raw MAP's could be attributed to individual differences in initial rest level.

The work-level MAP's for subclass conditions were found to contain heterogeneous error variance even after adjustment was made for initial rest differences. Bartlett's test gave a chi square of 29.03, significant beyond the .01 level. A cube-root transformation of the adjusted scores provided the homogeneity required for use of analysis of variance. Chi square (corrected for sample size) was then reduced to 4.83 ($P > .10$).

The results of the analysis of variance of the transformed adjusted work level scores are presented in Table 2. The *F* ratio for difficulty is significant beyond the .01 level. The interaction D × I is significant beyond the .001 level. The instructions variable, however, not only had an *F* ratio of zero, but led to a reversal, significant at the .05 level, with the mean-square value for instructions as the denominator of the *F* ratio. However, since adequate sampling and design precautions were taken, this is interpreted in accord with Edwards' views as being one of the "occasionally significant values which are to be expected in random sampling" (**11,** p. 228).

TABLE 2

ANALYSIS OF VARIANCE OF CUBE-ROOT WORK MAP's WITH INITIAL REST LEVELS REMOVED

Source	*df*	Mean Square	*F*
Difficulty	1	.7992	13.22
Instructions	1	.0001	.00
D × I	1	1.6921	28.00
Error	32	.0604	
Total	35*		

* Four degrees of freedom were lost in removing correlated rest measures. $N = 40$.

TABLE 3

ANALYSIS OF VARIANCE OF MEAN TIME (IN SECONDS) PER PROBLEM DURING SOLUTION

Source	*df*	Mean Square	*F*
Difficulty	1	8928.01	150.75**
Instructions	1	344.57	5.78*
D × I	1	90.60	1.52
Error	36	59.58	
Total	39		

* Significant at 5% level.
** Significant at .1% level.

Verbal responses.—Number of correct problems solved showed differences only between levels of difficulty. The totals for the four groups were as follows: E-E, 24; T-E, 26; E-H, 4; and T-H, 2. No further analysis of the accuracy data was made since the raw scores indicated no possible effect of the instructions variable.

Although the working time on the problems was artificially curtailed by the 60-sec. time limit, there nevertheless appeared to be differences in time scores among the groups. The mean time in seconds for each of the four groups was: E-E, 21.60; E-H, 48.56; T-E, 12.72; T-H, 45.70. Bartlett's test gave an uncorrected chi square of 4.93 ($P > .10$) indicating homogeneity of error variance. An analysis of variance (See Table 3) was therefore carried out.

DISCUSSION

It had been expected that during the instruction period proper the ego-orienting instructions would serve to arouse *S*s more than the task-orienting instructions. Such an anticipation followed from the observations made by other investigators (**2, 3, 19**) that ego-orienting instructions of the kind used in this

experiment seem to arouse a certain amount of anxiety. The present finding shows, however, that the two different kinds of instructions do not produce immediate differential changes in MAP solely as a function of being presented to *S*. The fact that work on problems following the instructions showed such effects in the interaction suggests that the differential effectiveness of these kinds of instructions on MAP does not appear until *S* is actually engaged in working on a task.

An examination of the nature of the interaction by comparing the means in the fourth column of Table 1, shows that the hypothesis of the experiment was substantiated. Ego-oriented *S*s do respond differently from task-oriented *S*s if the level of difficulty of the task is taken into account.

Previous studies by Hadley (**14**) and Shaw and Kline (**28**) have shown that increases in difficulty of problems are accompanied by increases in MAP. On the basis of these findings, these writers concur with Davis (**9, 10**) in considering the tension changes as an index of mental effort. If it can be assumed that these investigators employed traditional, task-orienting instructions,[4] then the findings of the present study are in essential agreement. Task-oriented *S*s exhibited higher MAP levels (and thus presumably more effort) on hard problems than they did on easy ones.

With ego-oriented *S*s, however, this relationship was reversed. These *S*s worked harder on the easy problems. Since both task- and ego-oriented *S*s were given the same problems, the difference between these groups must lie in the effects of the instructions. There are no available studies in which MAP differences are systematically related to differences in instructions of the kind employed here. Hence an understanding of the findings must be sought in terms of their relationship to the results obtained in related fields of investigation.

From the standpoint of studies of level of aspiration (**22**), task-oriented *S*s could be expected to work as hard as the problems demanded with minimal success-failure effects attached to their performance. In comparison, ego-oriented *S*s could be expected to display considerable effort on easy problems in order to make very certain that the induced goal of avoiding failure is achieved. Similar behavior was observed by Davis (**9**) and Clites (**6**). For difficult problems, ego-oriented *S*s might very well refuse to set a performance goal or adopt a private substitute goal—a reaction which would serve by default as a protection against failure (**5, 16, 20, 22**). In the latter instance, an index of mental effort should be low. This explanation is supported by the "criss-cross" interaction represented by the work-level means contained in Table 1.

Also applicable is Goldstein's "catastrophic reaction" hypothesis (**13**), especially in the case of reduced MAP levels for E-H *S*s. For difficult problems, *S*s could have regarded their inability to derive a solution as a personal defect and thereupon attempted to "withdraw" psychologically from the situation.

The fact that accuracy scores, as compared to MAP indices and time scores, were relatively insensitive measures of problem-solving activity confirms an earlier finding by Davis (**9**) and a more recent one by Adamson (**1**). Because of the 60-sec. time limit imposed for practical reasons on *S*s' working time, time-score differences were not fully analyzable. Nonetheless, the trends present in these data were consistent with the interpretation of the MAP measures. The longer time scores for E-E *S*s could reflect deliberation to make very certain that a solution is accurate. In the case of the E-H *S*s, increased time would probably represent deliberation before

[4] Only Shaw and Kline (**28**) give the exact wording with which their tasks were presented to *S*. Davis (**9**) describes his instructions, but does not give the exact wording. Both of these sets of instructions appear to the writer to be primarily task-orienting. Although Hadley (**14**) gives no description of his instructions, his procedural approach as well as the outcome of his experimental findings suggest that he, too, primarily task-oriented his *S*s.

deciding to "give up." Longer time measures for ego-oriented *S*s would then support either a level of aspiration or Goldstein's hypothesis.

Summary

To test the hypothesis that ego-instructed *S*s would have different MAP changes during problem solving than task-instructed *S*s if problem difficulty were controlled, 40 *S*s (equated for intelligence) were assigned 10 to each group of conditions in a two-by-two factorial design. The *S*s were given either five easy or five difficult problems in number progressions under either task- or ego-orienting instructions. MAP's were photographically recorded during initial (criterion 10μv or less) rest, administration of instructions, problem solution (work), and between-problem rest.

1. During the instructions period proper and during between-problem rest, no significant change of MAP over initial resting level was found.

2. The variance of transformed work-level scores, adjusted for correlation with rest level, was analyzed. Problem difficulty and the instructions-by-difficulty interaction had *F* ratios significant at the .01 and .001 levels, respectively.

3. The adjusted work MAP changes of the task-oriented *S*s confirmed earlier findings of Shaw and Kline (**28**) and Davis (**9**). The MAP changes of both the ego- and task-oriented *S*s are interpreted as being in agreement with both a level-of-aspiration hypothesis and with Goldstein's (**13**) "catastrophic reaction" hypothesis.

4. Time to solution of the problems tentatively supports either hypothesis, but accuracy scores showed only differences in difficulty of the problems. It was concluded that tension increase during problem solution was a function of the combined effect of the type of instructions and the difficulty of the problems used.

References

1. Adamson, R. E. Functional fixedness as related to problem solving: a repetition of three experiments. *J. exp. Psychol.*, 1952, **44**, 288–291.
2. Allport, G. W. The ego in contemporary psychology. *Psychol. Rev.*, 1943, **50**, 451–478.
3. Alper, T. G. Task-orientation vs. ego-orientation in learning and retention. *Amer. J. Psychol.*, 1946, **59**, 236–248.
4. Alper, T. G. Task-orientation and ego-orientation as factors in reminiscence. *J. exp. Psychol.*, 1948, **38**, 224–238.
5. Alper, T. G., & Black, A. The effects of instruction, task and population-sample on mental set. *Amer. J. Psychol.*, 1949, **62**, 295–299.
6. Cliters, M. S. Certain somatic activities in relation to successful and unsuccessful problem solving: III. Relation of action potentials and muscular movements and tensions to successful and unsuccessful problem solving. *J. exp. Psychol.*, 1936, **19**, 172–192.
7. Courts, F. A. Relationships between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, **25**, 235–256.
8. Davis, R. C. The relations of certain muscle action potentials to "mental work." *Indiana Univ. Publ. Sci. Ser.*, 1937, No. 5.
9. Davis, R. C. Patterns of muscular activity during "mental work" and their constancy. *J. exp. Psychol.*, 1939, **24**, 451–465.
10. Davis, R. C. Set and muscular tension. *Indiana Univ. Publ. Sci. Ser.*, 1940, No. 10.
11. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
12. Freeman, G. L. Changes in tonus during completed and interrupted mental work. *J. gen. Psychol.*, 1930, **4**, 309–334.
13. Goldstein, K. *The organism—a holistic approach to biology.* New York: American Book, 1939.
14. Hadley, J. M. Relationships between electric signs of central and peripheral activity during "mental work." *J. exp. Psychol.*, 1941, **28**, 53–62.
15. Holt, R. R. Effects of ego-involvement upon levels of aspiration. *Psychiatry*, 1945, **8**, 299–317.
16. Holt, R. R. Levels of aspiration, ambition or defense. *J. exp. Psychol.*, 1946, **36**, 398–416.
17. Hunt, J. McV. Psychological loss in paretics and schizophrenics. *Amer. J. Psychol.*, 1935, **47**, 458–463.
18. Jacobson, E. The electrophysiology of mental activities. *Amer. J. Psychol.*, 1932, **44**, 677–694.
19. Klein, G. S., & Schoenfeld, N. The influence of ego involvement on confidence. *J. abnorm. soc. Psychol.*, 1941, **36**, 249–258.
20. Lewis, H. B., & Franklin, M. An experimental study of the role of the ego in work: II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.

21. Lewin, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936.
22. Lewin, K., Dembo, T., Festinger, L., & Sears, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 333–378.
23. Marrow, A. J. Goal tensions and recall. *J. gen. Psychol.*, 1938, **19**, 301–344.
24. Max, L. W. An experimental study of the motor theory of consciousness. IV. Action current responses during awakening, kinaesthetic imagery and abstract thinking. *J. comp. Psychol.*, 1937, **24**, 301–344.
25. Osler, S. F. Intellectual performance as a function of two types of psychological stress. *J. exp. Psychol.*, 1954, **47**, 115–121.
26. Reuder, M. E. The relationship between two kinds of instruction and two levels of problem difficulty in their effects on muscular tension levels during problem solution. Unpublished doctoral dissertation, Univ. of Pennsylvania, 1951.
27. Shaw, W. A. The relation of muscular action potentials to imaginal weight lifting. *Arch. Psychol.*, 1940, No. 247.
28. Shaw, W. A., & Kline, L. H. A study of muscle action potentials during attempted solution by children of problems of increasing difficulty. *J. exp. Psychol.*, 1947, **37**, 146–158.
29. Smith, A. A. An electromyographic study of tension in interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 32–36.
30. Stauffacher, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26–46.
31. Wallen, R. Ego involvement as a determinant of selective forgetting. *J. abnorm. soc. Psychol.*, 1942, **37**, 20–39.

(Received March 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

RETROACTIVE INHIBITION AS A FUNCTION OF THE TEMPORAL POSITION OF THE INTERPOLATED LEARNING[1]

JOHN M. NEWTON AND DELOS D. WICKENS

The Ohio State University

All major theories of forgetting predict that the time relationship between original learning and interpolated learning will have some effect on the magnitude or retroactive inhibition (RI). For example, the perseveration theory of Muller and Pilzecker (**5**, pp. 432–437) predicts that temporal nearness of the interpolated learning (IL) to original learning (OL) should produce a decrement in retention. Likewise, because of forgetting of OL, the response strength of IL should be most equal to that of OL when the two occur close together in time. Thus a competition-of-response theory would also predict that this situation would yield the most RI (**5**, pp. 437–443). A two-factor theory, on the other hand, would state that in addition to response competition there is "unlearning" involved which accounts for much RI. Following Underwood's interpretation, this may be thought of as extinction of OL (**5**, pp. 444–447). It would thus be predicted that retention would be best when there is time enough after IL for recovery of OL to occur.

The evidence to date has been contradictory. It has been adequately summarized by McGeoch and Irion (**5**, p. 427) as follows: "The relationship between point of interpolation, length of interval, and other variables of this type of experimentation have not been sufficiently worked out to permit any generalized statement concerning point of interpolation."

In a recent paper, Archer and Underwood (**1**) also conclude that the relationship of the temporal point of IL to RI is still indeterminate. These *E*s systematically varied both point of interpolation and degree of IL in order to test for interaction between the two. While the degree of IL was found to be a significant variable, neither temporal point of IL nor the interaction had a significant effect on RI. There is, however, one possible variable that was not controlled in this experiment, and that might contribute to the experimental results. This is the phenomenon of "warm-up."

Irion (**2**) and Irion and Wham (**3**) have demonstrated that retention of verbal materials is facilitated by warm-up just previous to recall. Furthermore, Thune (**6**) has shown that paired-associate learning is about equally facilitated by both color guessing and paired-associate learning practice. The presentation of the IL just before the recall session can be thought of not only as a source of interference but also as a form of warm-up for that recall session. The groups whose IL precede recall by longer periods of time do not have this warm-up just before recall (though they do have the interference) and their performance may be depressed because of its lack.

Still another explanation of the Archer and Underwood results might lie in the high degree of RI that was intrinsic to their design. This was

[1] Experiment I in this paper is based on a thesis submitted by the first author in partial fulfillment of the requirements for the M.A. degree at The Ohio State University, 1952.

obtained by making the lists of the S_1—R_1, S_1—R_2 pattern, i.e., the interpolated list differed from OL only in the response words. While this has the advantage of making for neat interpretations in terms of either competition of responses or extinction of OL responses, it seems possible that the design may produce such large amounts of interference for all groups that it masks the interference which might be contributed by a variable of lesser significance.

This paper will present three experiments. These have been designed so as to duplicate as much as possible that part of the Archer and Underwood experiment which dealt with temporal point of IL, and thus to make results comparable. In Exp. I, the effect of introducing warm-up before recall on the variable of temporal point of IL is investigated. In Exp. II and III the effect of using relatively dissimilar material in IL is investigated in respect to the same variable.

Method

Equipment.—Two lists of 10 two-syllable adjectives were learned by each *S*. Experiments I and II differed from each other only in the materials used in IL. In Exp. I both lists were the same as those used by Archer and Underwood (1). These adjectives were selected originally on the basis of rules listed by Thune and Underwood (7) and follow the S_1—R_1, S_1—R_2 design. In Exp. II, one of the Archer and Underwood lists was retained and the second was modified so as to make both the stimulus and the response word different from those of the first list. Experiment III was essentially the same as Exp. II, except for a slightly different treatment of the IL task.

All lists were typed on white glazed-cloth tape and presented in a Patterson S-PA Memory Drum, Model I-A. Each stimulus word was presented alone for 2 sec., and with the response word for 2 sec. To discourage any serial effects, the lists were presented in three orders. The orders used by Archer and Underwood were followed, except for the modified list, in which case a sequence parallel to theirs was used. After each trial (one presentation of 10 pairs of words), there was an 8-sec. rest interval. The *S*s were run in a semi-soundproof booth and all distractions were minimized.

Subjects.—The *S*s were 180 students (60 in each experiment) enrolled in elementary psychology courses at The Ohio State University. All *S*s were volunteers for the experiment, but were serving to fulfill course requirements. No *S*s had previous experience with a memory drum. Forty-one of the *S*s were male and 139 female. Subjects of each sex were about equally divided among groups.

Procedure.—The usual RI design of learn List A, learn List B, test for retention of List A was used in all experiments. The first list was always learned to a criterion of one perfect recitation and was relearned 48 hr. later to the same criterion. In Exp. I and II the second list was learned to the same criterion at one of three time intervals after OL. These intervals were 0, 24, and 48 hr., respectively. About 1 min. elapsed between the two lists under the 0-hr. and the 48-hr. conditions. Experiment III employed the same intervals, but all *S*s were given 10 presentations of the IL list rather than learning it to a criterion.

In all experiments the 0- and 24-hr. groups were given warm-up trials immediately prior to relearning OL. The warm-up list was made up of paired numbers such that the response word was always the number which was one greater than the stimulus word. The numbers were written as words on the tape rather than printed as arabic numerals. The stimulus numbers ran from 1 to 10 and were randomly arranged. Three orders of arrangements were used. The list was presented in the same manner as the lists of adjectives and each *S* was given three trials of 10 pairs each. The *S*s were informed beforehand as to the method of pairing the numbers, and in no case did an *S* make any errors on the warm-up task. It was felt by *E* that this gave ample warm-up with a minimum of extraneous learning. No warm-up was given to the 48-hr. group, since for this group the IL should act as a warm-up task. In addition, *S*s in the 24- and 48-hr. IL groups of Exp. III were given six trials of this same warm-up task just previous to IL. To counterbalance any possible differential difficulty effects of the two lists, one list was used as the OL for half the *S*s and as the IL for the other half.

All *S*s received standard instructions for paired-associate learning. In Exp. I, *S*s were told of the relationship between OL and IL lists just previous to IL. In Exp. II and III, they were specifically told that there was no relationship whatsoever between the lists. Before relearning, effort was made to create a "set" to recall the first list by such methods as going over

the original instructions, etc. The *Ss* were also requested not to rehearse either of the lists, and were not informed of the nature of the experiment previous to completing it.

The relationship of the stimulus and response units for OL and IL in the two experiments is as follows:

Exp. I (OL) S_1—R_1; (IL) S_1—R_2
Exp. II and III (OL) S_1—R_1; (IL) S_2—R_2

Results

Original learning.—In each experiment a statistical check was run to indicate whether *Ss* in each of the six groups (three experimental conditions and two orders of presentation of the lists) were equal in initial learning ability, as indicated by the number of trials to learn the first list. In all three cases, Bartlett's test indicated that homogeneity of variance could safely be assumed. The total variance in each experiment was then analyzed. The *F* ratio was not significant at the .05 level of confidence in any case. Hence, we may assume that the groups in each study were reasonably equal in learning ability and that the lists used were of about the same difficulty.

As has been previously mentioned, the lists used in these experiments were the same as these of Archer and Underwood with the exception of one list in Exp. II and III. They were designed to be of equal difficulty, and this was borne out in the Archer and Underwood study. As a further check, however, in each condition of both experiments the lists were used equally often as OL and IL. The effect of order of presentation was then included in the analysis of variance done in each experiment for the two indices of RI used. In no case did this turn out to be a significant variable, which upholds Archer and Underwood's findings.

Table 1 presents a general summary of the experiments in terms of mean scores for all groups in all sessions. In view of the fact that no significant difference was found among the lists, the results for all lists have been combined. The table also includes the results of a control group of 10 *Ss* in which no interpolated learning was given, and it permits some judgment to be made on the absolute amount of RI that the experimental procedures generated.

TABLE 1

Mean Scores for All Experiments

Interval OL-IL	Trials for OL	Trials for IL	Recall	Trials to Relearn
Experiment I				
0 hr.	23.00	17.75	2.55	6.20
24 hr.	23.70	20.75	3.70	7.45
48 hr.	23.05	19.35	2.40	5.85
Experiment II				
0 hr.	20.00	10.15	5.85	4.55
24 hr.	20.00	16.10	4.85	5.45
48 hr.	23.30	21.20	3.25	4.75
Experiment III				
0 hr.	19.20		5.30	4.55
24 hr.	17.75		4.30	5.15
48 hr.	18.60		3.00	4.85
Control	21.70		6.40	4.00

Relearning.—The total variance of the relearning scores was analyzed for Exp. I and II. The *F* ratios obtained were 1.28 and 1.68, respectively, neither of which is statistically significant. According to Bartlett's test, the assumption of homogeneity of variance was not met for these scores in the third experiment. In view of the fact that relearning had not proved a satisfactory indicator of differences in RI previously, and since inspection of the means would indicate little likelihood of there being any signifi-

TABLE 2

ANALYSIS OF VARIANCE FOR RECALL SCORES: EXP. II AND III

Source	df	Experiment II		Experiment III	
		Mean Square	F	Mean Square	F
Between orders of presentation	1	2.82	—	5.40	1.66
Between treatments	2	34.40	9.77*	26.60	8.18*
Order × treatment	2	0.86	—	1.80	
Within groups	54	3.52		3.25	
Total	59				

* $P < .01$.

cant differences, no further statistical treatment was attempted.

Recall.—The recall scores of the groups within each of the three experiments were compared by means of analysis of variance after Bartlett's test showed, in each case, that the assumption of homogeneity of variance was met. In Exp. I, an over-all *F* test yielded a ratio of 1.43 which is not significant at the .05 level of confidence. In Exp. II, a double classification analysis was done to test the effects of the two orders of presentation and the three experimental conditions. This is presented in Table 2. The within-groups variance proves to be the correct error term. The *F* ratio for experimental treatment (temporal points of IL) is significant beyond the .01 level. Order of presentation does not prove to be an important variable, as would be predicted on the basis of the way that the lists were chosen. Since this was the case, the two orders were combined for each of the three treatments. This yielded three groups, each with an *N* of 20. Differences between the means of these groups were then tested by means of *t*. Comparisons between the 48-hr. group and the 0- or 24-hr. groups produced *t*'s of 4.44 and 2.88, respectively. In both cases the *P* values are less than .01. The difference between the 0- and 24-hr. groups produced a *t* of 1.62, which is not significant at the .05 level. The 48-hr. group has clearly poorer retention on recall scores than do the 0- or 24-hr. groups.

The scores of Exp. III were analyzed in the same manner as those of Exp. II. The analysis of variance is presented in Table 3. As in Exp. II, the *F* ratio for experimental treatment is significant beyond the .01 level, and order of presentation is not important. Again as in Exp. II the two orders were combined for the three treatments and differences were tested by means of *t*. Differences between the means of all three groups are significant beyond the .01 level of significance. Between the 0-hr. group and the 24- and 48-hr. groups the *t*'s are 5.87 and 12.42, respectively. Between the 24- and 48-hr. groups, *t* is equal to 7.08.

Intrusions.—In their study, Archer and Underwood also considered overt intrusions at recall as a possible index of RI. They define these as "the giving of a second-list response in place of a first-list response when *S* is trying to recall the first of the two lists" (1, p. 286). Following their example in this respect, the number of intrusions were tallied for each of the three conditions in Exp. I. These are as follows: 0-hr. IL, 13 intrusions; 24-hr. IL, 16 intrusions; and 48-hr. IL, 2 intrusions. These frequencies exhibit essentially the same trend as those obtained by Archer and Underwood, but are numerically much fewer than they report. Because of the low frequencies and questionable reliability of these data, no statistical treatment was attempted.

No intrusions at all from the interpolated list occurred during the relearning trials of Exp. II and III.

This is not surprising in view of the fact that the OL and IL were completely different in both stimulus and response words.

Discussion

The results of the Archer and Underwood experiment and our Exp. I (which, with the exception of our use of a warm-up task, was a repetition of theirs with twice the number of *S*s in the separate groups) are in complete accord in finding no relationship between temporal point of interpolation and degree of RI.

The results of Exp. II and III, however, suggest that this conclusion may be limited to the type of interlist relationship employed in Exp. I, namely, an S_1—R_1; S_1—R_2 relationship. This is a relationship which is designed to produce a maximum amount of RI, and the intensity of this condition may obscure the RI which can arise from a variable of lesser importance—as the temporal variable may well be.

In Exp. II and III, a positive relationship between temporal point of interpolation and RI was found. The direction of this relationship is, however, just the opposite of that which would be predicted by a perseveration theory. Although we have no positive evidence that our warm-up task did facilitate ability to recall, if we assume that it does do so from the work of Irion (2), and Irion and Wham (3), it is worth noting that a failure to include this procedure might produce results which spuriously favor the perseveration theory. It is possible also that in the present experiments the performance of the 0- and 24-hr. groups could have been depressed if no warm-up had been given, and the inferiority of the 48-hr. group might have been concealed.

We do not feel that at the present time it is possible to offer any clear theoretical formulation for the results obtained in Exp. II and III. Because of the fact that our interlist relationship was an S_1—R_1; S_2—R_2 relationship it is not easy to fit the finding into the two-factor theory. This does not deny the possibility that unlearning and response competition may have operated despite stimulus and response dissimilarity. They may have done so to a minor degree through the mechanisms of stimulus and response generalization, but in ways that are too obscure to be easily identified. Apropos, it should be recalled that no intrusion occurred in Exp. II or III.

There is a possibility that the results may have arisen from a more generalized form of competition than is usually understood by the competition-of-response theory. Ordinarily one thinks of competition of response in reference to a particular stimulus. It is possible, however, that *S*s may acquire a tendency to give a certain class of responses in addition to giving specific responses to specific stimuli. The class of response referred to would be, of course, any of the responses employed in a given list. This generalized tendency—a kind of set—would be strongest for the 48-hr. group since it must make its recall of the R_1 class immediately after the learning of the R_2 class. Hence their recall scores might be depressed. Further work is now in progress testing this hypothesis.

One possible artifactitious type of explanation may be found in the data of Exp. II. Here the mean number of trials required to learn the interpolated list is greatest for the 48-hr. group, and Archer and Underwood had found that the amount of RI is positively related to the degree of IL. Thus on the basis of their experiment it might be argued that the inferiority in recall of the 48-hr. group arose from the large number of trials that this group required in IL. The argument, however, is not strictly justified, for the variable with which they worked was the *degree* of learning, and this was controlled in Exp. II, for all groups learned to the identical criterion, though the number of trials required to achieve this criterion varied among the groups.

As an additional control, Exp. III was run. It will be recalled that this differed from Experiment II in that all *S*s received 10 trials on the IL list. The results of the two studies were identical. If any-

thing, the latter was more conclusive. It is worthy of note, however, that the same relationship between time of interpolation and speed of learning was found in both experiments. As an indication of this, the mean number of correct responses on the tenth IL trial was computed for each group. For Groups 0, 24, and 48, these turn out to be 8.7, 7.1, and 6.9, respectively. The *t* test shows that Group 0 differs from the other two beyond the .01 level of confidence, although no significance is shown for the difference between Groups 24 and 48. Here we have the rather amazing situation of greater RI associated with lower degrees of interpolated learning. These differences in degree of IL might be attributable to differential warm-up, but the likelihood of this is diminished by the fact that a rather extensive warm-up task was provided for Groups 24 and 48 previous to IL. Of course, the warm-up of learning an actual list might be far superior to the artificial warm-up provided Groups 24 and 48. It this were so, however, one would expect that the relearning trials for Group 48 would be materially less than for the other groups.

The findings of Exp. II and III seem to be confirmed by a similar study by Maeda (4). Although in this latter experiment much shorter time intervals between OL and IL were used, essentially the same results were obtained. It is difficult to determine the exact nature of the learning materials from the published account of this study, but OL and IL were dissimilar verbal tasks.

Summary

Three experiments were conducted to investigate the relationship between degree of retroactive inhibition and point of interpolation of the interfering materials.

A 10-unit list of paired-associate adjectives was learned to a criterion of one perfect reproduction. In Exp. I and II an interpolated list of the same type was learned to the same criterion 0, 24, or 48 hr. afterwards. In Exp. III the same intervals were used, but each *S* was given 10 trials on the IL list, rather than learning it to a criterion. Recall occurred 48 hr. after the original learning. The 0- and 24-hr. groups were given a warm-up task before recall. In Exp. I the relationship between the two lists was S_1—R_1; S_1—R_2. In Exp. II and III the relationship was S_1—R_1; S_2—R_2.

No relationship between point of interpolation and magnitude of RI was found in Exp. I. In Exp. II and III such a relationship was found, with the 48-hr. group showing significantly more RI than either of the others. The findings were significant for recall but not for relearning scores.

References

1. Archer, E. J., & Underwood, B. J. Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *J. exp. Psychol.*, 1951, **42**, 283–290.
2. Irion, A. L. Retention and warming-up effects in paired-associate learning. *J. exp. Psychol.*, 1949, **39**, 669–675.
3. Irion, A. L., & Wham, D. S. Recovery from retention loss as a function of amount of pre-recall warm-up. *J. exp. Psychol.*, 1951, **41**, 242–247.
4. Maeda, Y. Zur experimentellen Untersuchung über Faktoren der Reproduktion. *Jap. J. Psychol.*, 1951, **21** (3–4), 1–17.
5. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
6. Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423–438.
7. Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of the degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–201.

(Received February 20, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

SPONTANEOUS RECOVERY OF R_1 FOLLOWING INTERPOLATED ACQUISITION AND EXTINCTION OF R_2[1]

ALBERT E. HICKEY, JR.[2]

Boston University

Spontaneous recovery is said to occur if the following sequence of behavior events can be demonstrated: (*a*) the acquisition of an S-R association, (*b*) some degree of its experimental extinction, (*c*) an interval during which the nominal CS is not presented, and (*d*) recurrence of the CS, with reinstatement of the originally associated response at greater strength than at the conclusion of the extinction phase (*b*). Experimental investigation of spontaneous recovery can be directed at any phase of this cycle. In this study the interval (*c*) between extinction and the test of spontaneous recovery was the focus of interest. Specifically, if responses *x* and *y* are learned and *x* is partly extinguished, what are the relative effects on spontaneous recovery of *x* of each of the following interpolated conditions: extinction trials on *y*, additional acquisition trials on *y*, and a comparable time interval involving neither acquisition nor extinction trials on *x* or *y*?

Liberman (**5**) performed two pioneer experiments which fit the foregoing paradigm. In both experiments the recovery of the galvanic skin response (GSR), conditioned to a tone, was measured as a function of the following activities: additional acquisition of the conditioned eyelid response, extinction of the conditioned eyelid response, and a comparable period of rest. He points out, however, that the GSR on the final trials of first extinction was "so low in magnitude and ambiguous in form" that these trials could not provide "a meaningful base-line for the calculation of spontaneous recovery" (**5**, p. 295). Therefore, in the determination of both relative and absolute recovery, Liberman was forced to assume that the terminal extinction levels for all three groups were not significantly different.

The purpose of the present experiment was to provide data on the recovery of an instrumental response, less labile than the GSR, in a design similar to that used by Liberman. White rats acquired both a bar-pressing and a panel-pushing response. Between the first and second extinction of the bar-pressing response each *S* received one of three different experimental treatments: additional acquisition trials on panel-pushing, extinction trials on panel-pushing, and a comparable period of rest.

METHOD

Subjects.—The *Ss* were 122 experimentally naive, male albino rats. Of these 93 were used in the main body of the experiment and 29 were used in a transfer test of task similarity. All *Ss* were between 70 and 90 days of age at the beginning of training and their average weight was 176 gm.

[1] This paper is based on a portion of a dissertation submitted to the Graduate School of Boston University in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express grateful acknowledgement to Dr. Leo J. Reyna for his direction of this investigation, and to Dr. Garry Margolius for many helpful suggestions.

[2] Now with the Electric Boat Division, General Dynamics Corporation.

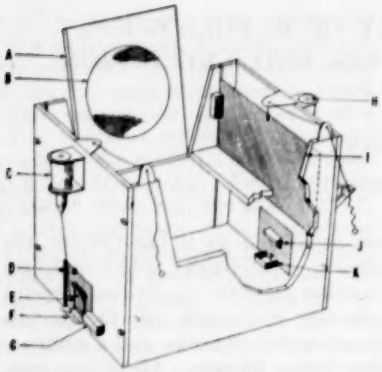

FIG. 1. Cutaway of the experimental box with bar (*J*) in extended position. The component parts are described in the text.

Apparatus.—The experimental box of unpainted plywood, shown in Fig. 1, was 11 in. long, 6 in. wide, and 5 in. high (inside dimensions). A movable bar (*J*) and a swinging panel (*E*), both of brass, were located in opposite ends of the box and each could be concealed by an aluminum shutter (*I*). The panel, 2½ in. by 1½ in., was hinged at the top to swing freely when a light pressure was applied from within the box. A ¼-in. displacement of the panel closed a BF2RW-22 Microswitch (spring removed) (*G*). A 6-v. solenoid (*C*) connected to the panel by a rubber cord opened the panel an additional ¾ in. disclosing to *S* a brass food trough (*F*).

Through a rectangular aperture 1½ in. by 1 in. in the opposite end wall projected a hollow brass prism (*J*) approximately triangular in cross section. The prism was hinged and balanced so that it normally extended ¾ in. into the box. The top surface of the prism in this position formed a small shelf or bar. A 4-gm. pressure depressed the bar ¼ in., closing a V4-14 Microswitch. The action of a solenoid (*H*), connected by a linkage to an external projection of the prism, completed the rotation of the prism so that the flat top surface closed flush with the face of the end wall. With the bar fully retracted *S* could make no additional bar-pressing responses. Food pellets, placed singly in a tube by *E* prior to each trial, dropped into a small brass food cup (*K*) directly below the bar when released by a trigger actuated by the same solenoid that retracted the bar.

Opening either shutter started a Standard Electric Timer which was stopped by the operation of the bar or panel. The light source was a 60-w. fluorescent lamp directly over and 8 in. above a 4½-in. diameter screened window (*B*) cut in door (*A*) on the top of the box. In addition to the experimental box there were three boxes for habituation and two for preliminary training, all of the same form and materials as the experimental box. The two preliminary boxes each contained a food cup on one end wall supplied through a tube projecting outside the box. The reinforcing pellets were tablets of Purina Lab Chow weighing .045 gm. with a diameter and thickness of 4 mm. The experimental room was darkened and sound dampened and internal noise was masked by the hum of an exhaust fan.

Procedure.—A one- to four-day period of ad libitum feeding intervened between receipt of *Ss* and the beginning of a 24-hr. feeding schedule. Habituation to the schedule extended from Day 1 to Day 5, during which *Ss* were removed daily from the living cages and placed for 1½ hr. in individual feeding cages, each containing 12 gm. of regular Lab Chow. There was unlimited access to water.

Preliminary training, to accustom *Ss* to eating pellets from the food trays, extended from Day 6 through Day 8. On Day 9 the 93 *Ss* in the main experiment received 10 acquisition trials on bar-pressing followed by 10 acquisition trials on panel-pushing. At its scheduled feeding time each *S* was placed in the experimental box with both shutters closed. After a 1½-min. adaptation period the shutter concealing the CS was presented by raising the shutter concealing the extended bar. The shutter was raised without regard to *S*'s orientation in the box. If *S* failed to operate the bar in 2 min., it was operated by *E* when *S* was in a position approximating that required to depress the bar. Since *S* performed only the consummatory part of such an assisted response, it was not counted as an acquisition trial and the procedure was repeated until 10 consecutive and unassisted responses, each within the 2-min. limit, were made. Following a 15-sec. interval allowed for pellet consumption, the shutter was closed for an additional 15-sec. period before starting the next trial.

Following the tenth bar-pressing response, 1 min. elapsed before the first presentation of the panel stimulus. The time schedule for the acquisition of panel-pushing was the same as that for the acquisition of bar-pressing. After completing Trial 10, each *S* was removed to its living cage for 15 min. before being placed in a feeding cage for 1 hr. of supplementary feeding on 10 gm. of regular Lab Chow.

On Day 10, 22 hr. after completing the two acquisition series, the 93 *Ss* underwent a 10-trial extinction on bar-pressing. The extinction procedure differed from the acquisition procedure only in that reinforcement was withheld and the

time allowed for response on each trial was limited to 30 sec. The 30-sec. stimulus presentation was introduced to control intertrial spontaneous recovery. On completing Trial 10 of this first extinction, *Ss* were assigned to three groups (labeled A, E, and R) such that the median latencies of these groups were as nearly equal as possible.

Group A received 10 additional acquisition trials on panel-pushing 1 min. after the shutter closed on Trial 10 of the first extinction. Group E received 10 extinction trials. Group R remained in the lighted box for 10 min. beginning when the shutter closed on Trial 10 of the first extinction. The three groups then concluded the experiment with a second series of 10 extinction trials on bar-pressing directly following the interpolated activity.

The remaining 29 *Ss* were used in a transfer test of task similarity. The preliminary training for this group was identical with that of *Ss* in the main experiment. However, they acquired bar-pressing after panel-pushing on Day 9, and then extinguished the two responses in the same order on Day 10. Comparison of this group with *Ss* in the main body of the experiment, which acquired and extinguished bar-pressing first, showed only positive transfer from the acquisition of bar-pressing to the acquisition of panel-pushing to be statistically significant. There was no significant transfer in the opposite direction nor was there significant transfer from the extinction of one task to the extinction of the other. Moreover, on observation, the typical bar-pressing and panel-pushing responses were quite dissimilar. The stereotyped bar response was a stroke of *S*'s paw while *S* characteristically used its nose to push open the panel.

Results

The acquisition of bar-pressing and panel-pushing followed the typical negatively accelerated decay function obtained when the median response latency is used as the group performance measure. The course of first and second extinction of bar-pressing

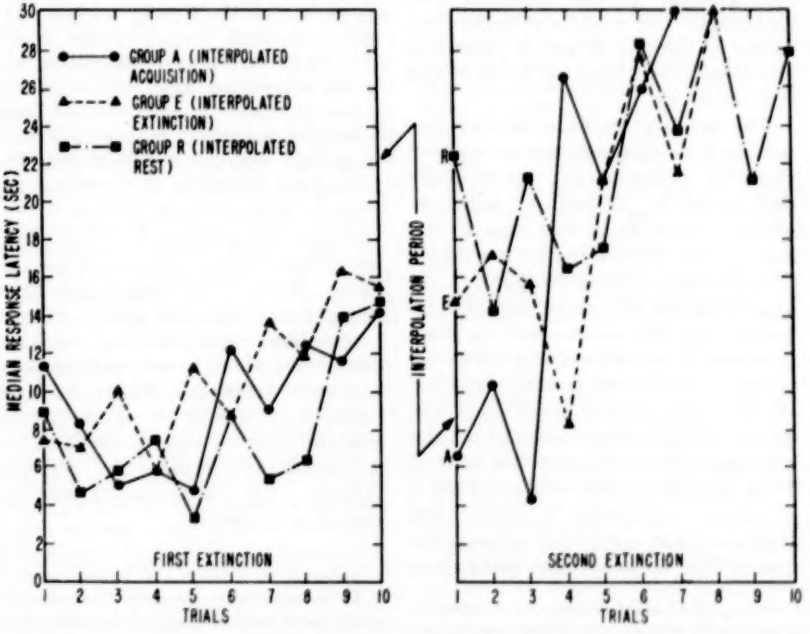

FIG. 2. First and second extinction of the bar-pressing response. During the interpolation period Group A received 10 additional acquisition trials on panel-pushing, Group E received 10 extinction trials on panel-pushing, and Group R received 10 min. of rest. *N* for each group was 31.

is plotted individually for each of the experimental groups, A, E, and R, in Fig. 2. The space separating the two parts of the figure represents the recovery interval. On first extinction all three curves show a drop in response latency between Trials 1 and 2, suggestive of warm-up, and thereafter follow a positive growth function of the number of trials. On Trial 1 of second extinction, following the interpolated conditions, the three groups are much more widely separated on the ordinate than they are on Trial 10 of first extinction.[3] The median response latency of Group A is shorter than that of Group E, which is, in turn, shorter than that of Group R. For the balance of second extinction the groups, despite increased variability, appear to follow the positive growth function noted in first extinction, the median response latencies of Groups A and E exceeding the 30-sec. trial limit on Trials 8 and 9, respectively.

The median response latency was used in drawing the curves since many stimulus presentations, particularly during second extinction, failed to elicit a response within the 30-sec. limit. Faced with similar truncated distributions, other investigators have approximated the mean and *SD* by using only the measurable response latencies (**2**) or by assigning the cutoff value to the indeterminate responses (**6**). In the present experiment the mean and *SD* were estimated for the truncated distributions of the critical Trial 10 of first extinction and Trial 1 of second extinction by employing Halperin's (**3**) method of approximation. The distributions were first normalized by converting the individual latencies to their logarithmic equivalents. The estimated means and *SD*'s for the two trials are shown in Table 1.

[3] Only Trial 1 of the 10-trial second extinction is used in the measurement of recovery since on any subsequent trial the three groups could be affected differentially by their performance on the first.

TABLE 1

MEAN LOG LATENCIES ON TRIAL 10 OF FIRST EXTINCTION AND TRIAL 1 OF SECOND EXTINCTION

Group	Trial 10		Trial 1		Critical Ratio		
	Mean	*SD*	Mean	*SD*	T_1		T_1-T_{10}
A	1.10	.63	.81	.37	3.252	5.053	2.229
E	1.20	.63	1.15	.46	2.268		0.376
R	1.11	.59	1.46	.62			2.325

The CR is used to estimate the significance of the differences among the means on Trials 10 and 1. The largest difference among the means on Trial 10 of first extinction could have occurred more than 50 times in 100 due to chance alone, supporting the pre-interpolation equivalence of the three groups. As with the medians, a comparison of the means on Trial 1 of second extinction shows that, following interpolation, Group R has the longest latency and Group A the shortest, while Group E is intermediate. The differences are all significant. Table 1 also makes possible a comparison, for each group, of pre-interpolation mean latency on Trial 10 of first extinction and postinterpolation mean latency on Trial 1 of second extinction. Group A shows a significant increase and Group R shows a significant decrease in response latency. No significant change occurred in Group E.

DISCUSSION

It is generally thought that an extinguished response will demonstrate an increase in response strength following a period of rest. Munn (**7**, p. 243), for example, reports that he found ". . . no

reported instance of failure to observe spontaneous recovery where it has been looked for soon after extinction." In addition to this empirical evidence there is Hull's (**4**, p. 284) theoretical position that reactive inhibition accrued during the extinction procedure dissipates as a function of time alone leading to spontaneous recovery of the extinguished response during rest. Liberman (**5**) also holds that activity during the interpolation interval, and presumably the random activity of the resting *S*, retroactively inhibits the null response acquired during the extinction procedure, leading to the apparent recovery of the original positive response.

In the present experiment, however, the rest group demonstrated a decline in the strength of the bar response during the interpolation period. Under the conditions of this experiment the correlated increase in response latency might fall under such a notion as loss of set-to-perform, a concept proposed by Ammons (**1**) in another context. Instead of being withdrawn from the experimental situation during the rest interval and replaced just before the second extinction, *S*s were confined to the box for the entire 10-min. period. During this confinement *S*s' expectancy or set-to-perform could have been lost. A simple cueing stimulus just before extinction might have restored the set. Some support for this hypothesis can be found in the abrupt drop in response latency in the rest group between Trials 1 and 2 of second extinction, a characteristic of warm-up not shared by the other groups.

In the studies of spontaneous recovery, the rest condition, largely because of single-habit *S*s, has received more attention than interpolated acquisition or extinction. With the exception of Liberman's study, no data appear in the literature to suggest the order of recovery of response following the interpolated acquisition or extinction of another response. This leaves deductions from the theories of Hull and Liberman as the strongest determinants of our expectations with regard to relative recovery. The two theories are in agreement that interpolated acquisition should lead to more recovery, and interpolated extinction to less recovery, than a comparable period of rest. In the present experiment, therefore, the greater recovery of the acquisition group relative to the extinction and rest groups is as expected. The fact that the rest group showed less recovery than the extinction group, however, does not conform to expectation. Supporters of Hull and Liberman might well argue that if the experimental conditions leading to loss of set-to-perform had not obtained for the rest group, the order of recovery for the three groups might have been consistent with both theories. The extinction group could be said to have retained a set-to-perform the bar-pressing task by performing the panel-pushing task even though it was unrewarded. The rest group, lacking interpolated activity, lost the set-to-perform. The implication that the performance of a nonrewarded response is, under certain conditions, more effective in maintaining the strength of a second response than a comparable period of rest should be explored in future multihabit experiments.

Summary

On the day following the acquisition of both bar-pressing and panel-pushing responses, 93 rats received two 10-trial extinctions on bar-pressing. Between the two extinction series, 31 *S*s received 10 additional acquisition trials on panel-pushing, 31 received 10 extinction trials on panel-pushing, and 31 received 10 min. of rest.

Between the final trial of first extinction, on which the three groups reached a common response level, and the first trial of second extinction, the interpolated acquisition group showed a decrease in bar-response latency, the interpolated rest group showed an increase in latency, and the extinction group showed no significant change. The relatively greater spontaneous recovery of the acquisition group conformed to theoretical expectation. The decline in response strength of the rest group during the recovery period, however, was theoretically and empirically unexpected, and resulted in the rest group demonstrating relatively less recovery than the extinction group. The conditions of the experiment suggest a loss during interpolated rest of set-to-perform.

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Graham, C. H., & Gagné, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.
3. Halperin, M. Estimation in the truncated normal distribution. *J. Amer. Statist. Ass.*, 1952, **47**, 457–465.
4. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
5. Liberman, A. M. The effect of interpolated activity on spontaneous recovery from experimental extinction. *J. exp. Psychol.*, 1944, **34**, 282–301.
6. Mote, F. A., & Finger, F. W. Exploratory drive and secondary reinforcement in the acquisition and extinction of a simple running response. *J. exp. Psychol.*, 1942, **32**, 57–68.
7. Munn, N. L. *Handbook of psychological research on the rat.* New York: Houghton Mifflin, 1950.

(Received March 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 2, 1956

INDIVIDUAL BEHAVIOR IN SOCIAL SITUATIONS: ITS RELATION TO ANXIETY, NEUROTICISM, AND GROUP SOLIDARITY [1]

VLADIMIR CERVIN [2]

University of Toronto

The study reported here describes an attempt to apply S-R theory to the prediction of the behavior of individuals in certain group situations. The problem investigated is the relation to behavior of three variables—anxiety and neuroticism (personality variables), and group solidarity (social variable). These variables are operationally defined and equivalence is established and tested between them and what are believed to be the corresponding learning variables—drive, habit strength produced by stimulus generalization, and sign (±) of secondary reinforcement, respectively. Behavior is measured by response latency, degree of participation, etc. Predictions are derived from learning theory in terms of reaction potential, and experimentally tested on individual behavior in controlled social situations.

1. It has been shown that the hypothetical construct of anxiety, interpreted as level of emotional responsiveness, is related to performance in conditioning (**11**), and that the difference in strength of performance between high- and low-anxious *Ss* is larger under an aversive stimulus (e.g., shock) than under a more neutral stimulus, such as air puff (**12**). This is in accordance with S-R theory, if it is assumed that anxiety and aversive stimulus each affect drive level (*D*), and that they combine in some way, e.g., multiplicatively. This combination will not only have the effect of increasing drive for all *Ss*, as compared with drive level produced by anxiety alone, but of increasing it more for high- than for low-anxious *Ss* (if the increase is proportional). Since drive is directly related to strength of performance, this explains the findings.

On the other hand, it has been shown that in a simple rewarding situation (**2**) no significant differences appear between high- and low-anxious *Ss*; it is also known that an accepting attitude of a therapist reduces anxiety in patients. If it is assumed that, on account of anxiety reduction, a proportional decrease of drive occurs in such situations, in contrast to proportional increase of drive in aversive situations, then not only should drive of all *Ss* be reduced, but it should be reduced more for high- than for low-anxious *Ss*. This is the first problem of the present study: the relation of anxiety to behavior under aversive and rewarding conditions.

2. Behavior which is invariant in time, or from situation to situation, has been observed in animals and humans. In animals it is the familiar response fixation (**4, 9**), interpretations of which range from the hypothesis that the amount of variability is a

[1] Research done under National Health Grant from the Government of Canada to Professor J. D. Ketchum.

[2] I am indebted to Professor J. D. Ketchum for advice and cooperation in carrying out this project. Thanks are also due to M. Humphries, N. Fletcher, and R. Gibbins for assistance in carrying out the experimentation, to Dr. A. H. Shephard and R. Teghtsoonian for criticizing parts of the manuscript, and to Joan Cervin for editorial help and typing.

function of the amount of motivation (**4**), to the compulsiveness hypothesis (**9**, p. 57). In human *Ss* invariant behavior appears in the form of widely generalized, maladjustive instrumental responses, e.g., compulsions, rationalizations. These reaction tendencies presumably represent defenses against anxiety which they reduce. Their clinical interpretation is somewhat similar to the motivation hypothesis (**4**).

However, it can be predicted from the S-R theory that an overlearned response tendency, such as a defense, should have a high habit strength and thus interfere with normally evoked responses in many situations. This interference may be expressed as withdrawal of generalized habit strength ($\bar{H}$) of defenses from the habit strength (H) of a normally evoked response: ($H - \bar{H}$) (**7**). Interference should occur when an individual does not discriminate a given situation from the original traumatic situation. The hypothetical construct of neuroticism (**5**, p. 49), which is inferred from defense responses, can be represented by the habit strength ($\bar{H}$). If habit strength is related to strength of performance, neuroticism should, under certain conditions, unfavorably affect performance. This is the second problem of the present study: the relation of neuroticism to behavior under two experimental conditions—one more and one less likely to evoke defenses.

3. Our recent findings (**3**) and those of others (**10**) indicate that performance of all *Ss* is better and more even under rewarding than under aversive conditions. This result can be predicted from behavior theory assuming certain values for the variables involved. But in social situations prediction of behavior and its testing is very difficult because of the constant response interaction between participants. Using one *S* with role-playing assistants who pose as *Ss*, however, one can devise social laboratory conditions analogous to those in learning and conditioning experiments: aversive stimuli (shock) can be produced and kept constant by aversive behavior of assistants, e.g., their opposition to *S*; positive (rewarding) stimuli, e.g., food, can be replicated by assistants' approving or accepting behavior. Aversive and approving behavior of "other" individuals in social situations operationally defines the hypothetical construct of solidarity of group members (**3**). Solidarity proper is conceptualized as mutual giving of positive, and dissolidarity of negative, secondary reinforcement by individuals in social situations. The number of reinforcements (N) and their signs thus can be controlled with the aid of assistants. The third problem of our study is the relation of solidarity to performance.

To test predictions stated under "Method," *Ss* with extreme scores on anxiety and neuroticism were subdivided into four categories—high-high, high-low, low-low, and low-high. Each of them was put through (*a*) a debate session in which his opinion was uncompromizingly opposed (aversive situation, expected to elicit *S*'s defenses, if any), and (*b*) a situation in which all his task-relevant suggestions were approved (rewarding situation, expected not to elicit defenses).

Method

Design of experiment.—A 2^3 factorial experiment was designed to vary the two personality variables and the one social variable (Table 1). Groups of three were formed, each with one *S* and two role-playing assistants of the same sex, unknown to him, who posed as *Ss*.

The solidarity situation was conceptualized as follows: *S*'s dominant response tendency (DRT), i.e., his own opinion, received reinforcement, and his defense reactions, if evoked, received extinction treatment. In the dissolidarity situation *S*'s DRT received extinction treatment, the nondominant response tendency (NRT), i.e., the

TABLE 1

EXPERIMENTAL DESIGN OF A SOCIAL SITUATION: THE SAME *Ss* AND DESIGN USED IN SOLIDARITY AND DISSOLIDARITY SITUATIONS

($N = 64$)

	Anxiety (*D*)			Mean *Pd* Score
	Low *Ss*	High *Ss*	Difference	
Neuroticism ($\bar{H}$)				
High *Ss*	16	16		10.4
Low *Ss*	16	16		16.8
Difference				6.4*
Mean Taylor Score	19.4	34.9	15.5*	

* Significant at the .001 level of confidence (1 tail).

opposite opinion, received reinforcement, as did also his defenses. The initial habit strength (H) of *Ss*' DRT was equalized, as far as possible, by attitude premeasurement and selection of topics important to them. Since all *Ss* were clinically normal, it was assumed that the habit strength (H) of DRT was larger than that of defenses ($\bar{H}$). Within each situation the number of reinforcements (N) and their magnitude (K) were kept constant by controlling the length of sessions and the behavior of assistants.

The strength (E) of response tendency was assumed to be a direct function of the intervening variables: $E = f(D, \underline{H}, K)$, where $\underline{H} = (H - \bar{H})$; E is therefore an inverse function of neuroticism ($\bar{H}$). The strength of response was assumed to be function of E and was measured by means of response latency (L), degree of participation (PQ), and frequency (R) of shift of opinions in the dissolidarity situation, as described below.

Subjects and personality variables.—About 500 students in two psychology classes were screened by several tests, of which only the Taylor Anxiety Scale and the *Pd* Scale—which satisfied the requirement of low (.07) intercorrelation—were retained as measures of anxiety and neuroticism, respectively. An inverted *Pd* Scale was used tentatively to measure neuroticism in the absence of better measures, on the assumption that the opposite of invariant behavior is the labile, impulsive behavior, such as that of psychopaths (characterized by lack of interference of controlling response tendencies), which is measured by the *Pd* Scale.[3]

[3] The assumptions made in the derivation of hypotheses are based on the behavior tendencies measured by the *Pd* Scale. Evidence that *Pd* and *Pt* items from MMPI form the opposite ends of a Guttman-type scale will be presented in a separate paper.

The number of *Ss* in each cell (Table 1) was randomly limited to 16, i.e., four for each sex and each course year, which gave a total of 64. Some very extreme cases, known to have had psychiatric backgrounds, were excluded. All selected *Ss* were English-speaking and most were Canadian. The dividing scores between high and low on each test were the means of the total test distributions: 27 (Taylor) and 13 (*Pd*).

Social variable.—In the solidarity situation each group of three was shown a set of six large TAT cards and asked to choose one picture about which to make up a story. They were then instructed to proceed in three phases: describe (*a*) the contents of the picture, (*b*) what was happening in it, and (*c*) what would happen. The assistants had had previous training and had been instructed how to behave: "Let *S* choose the picture . . . When *S* does make a suggestion then, after some mild argument . . . warmly approve the suggestion . . ." (operational definition of solidarity). "*S* should feel that here he has really made a success . . . ." Besides giving *S* encouragement, the procedure also aimed at obtaining, through the use of the projective possibilities of the situation, his opinion on a topic of some importance to him. Groups were told that they would be watched and timed. Mean duration of sessions, without *S*'s latency time, was 5 min., 20 sec. This time was empirically established and subsequently all sessions were kept down to it as far as possible, but without actually cutting them short.

In the dissolidarity situation each member of the same groups of three who participated in the solidarity situation was instructed to fill out a questionnaire[4] stating his opinions (pro or con) and their strength (importance to himself) on several topics chosen from a survey of students' interests. The questionnaires were collected and, during another task, the assistants' questionnaires were altered by *E* so that their opinions conflicted with those *S* marked "very important." A card was drawn to choose at random one "very important" topic for discussion. All questionnaires were then returned to their writers and the group instructed to debate the chosen topic with the aim of reaching consensus of opinion. "If no common opinion can be worked out, the group may decide to disagree." Assistants' instructions were: "Take the initiative in expressing your opinion from the word 'go' . . . this opinion will be the opposite of that of *S*" (operational definition of dissolidarity). "If *S* agrees with you, suggest closing the discussion and signaling agreement."

[4] The reason for giving entirely different tasks in the two situations is that similar, but reversed, tasks given *Ss* in pretrials simply killed their response by making them suspicious of the authenticity of other "*Ss*."

The order of the situations was randomized by tossing a coin. To control the length of sessions, the observer, after an empirically established time of 8 min. from the beginning of the discussion, suggested to the group that they arrive at a decision. Mean duration of sessions, without *S*'s latency time, was 7 min.

Behavior measurements.—All sessions took place in a demonstration room separated from an observation room by a one-way vision screen. Voices could be transmitted both ways between the two rooms. There were two observers. Discussions were completely recorded on tape. Behavior was measured as follows:

1. *Response latency*, i.e., the time between the word "go" and the moment when *S* made his first statement, was measured by stop watch and recorded. In the solidarity situation, *S*s' latency in the third phase was taken as the measure comparable to that in the dissolidarity situation. Reliability checked against latencies obtained on the "Interaction Chronograph" operated by a different observer was better than .99.

2. *Opinion latency*, i.e., the time between the start and the moment when *S* first stated his opinion, was measured and recorded. Since sometimes there was no clear-cut statement of opinion, the measure is incomplete and unreliable; results will not be given.

3. *Degree of participation* (PQ) in discussions was measured as the quotient of time *S* spoke to the total duration of the situation. Data for this measure were obtained from the "Interaction Chronograph" on which each speech of every person was recorded; the lengths of these speeches were then summed up and divided into the total length of the session. Reliability against recheck was over .99.

4. *Shift of opinion* was recorded at the end of the debate by the observer as "complete shift," "qualified shift," and "no shift." No errors were found on recheck.

In each session the same *E* read the instructions, the same experimental material was used, and with a few exceptions, the same pairs of assistants worked together. Some other measures were taken which await analysis: ratings of *S*s' "involvement" in the tasks, voice levels during discussions, and many qualitative observations.

Derivation of predictions.—These were derived from assumptions and theory by a little simple algebra of inequalities. Predictions of the degree of participation (PQ), which is a direct function of *E*, were derived as follows:

1. In the solidarity situation, high- and low-neurotic *S*s should have equal habit strength (H) of DRT and drive (D), and little or no defense interference ($\bar{H} = 0$); high-anxious *S*s should have the same habit strength (H) as low-anxious *S*s, but a somewhat higher drive (D), in spite of its greater reduction. Prediction: more or less equal E for all *S*s, and therefore also equal PQ.

2. In the dissolidarity situation, low-neurotic *S*s should have the same drive (D) as high-neurotic *S*s, but a higher habit strength (H) because of low strength ($\bar{H}$) of interfering response tendencies: $(H - \bar{H}_{high}) < (H - \bar{H}_{low})$; high-anxious *S*s should have the same habit strength (H) as low-anxious *S*s, but a higher drive (D). Prediction: higher E for low-neurotic and high-anxious *S*s as compared with their respective counterparts, and therefore higher PQ.

3. Comparing the predictions for the two situations, one would expect a smaller total increase of PQ from dissolidarity to solidarity for high-anxious and low-neurotic *S*s as compared with low-anxious and high-neurotic *S*s, respectively (interaction effect).

Predictions of response latency (L), which is an inverse function of E, were derived as follows:

1. In the solidarity situation predictions follow the same pattern as those of PQ because of the assumed absence of interfering defenses.

2. In the dissolidarity situation (*a*) E of high-neurotic *S*s should be higher, and therefore latency of response shorter, than that of low-neurotic *S*s because, D being constant, the combined[6] habit strength $(H + \bar{H})$ of this response is larger for high- than for low-neurotic *S*s: $(H + \bar{H}_{high}) > (H + \bar{H}_{low})$. (*b*) With H constant, E of high-anxious *S*s should be higher, and therefore latency shorter, than that of low-anxious *S*s.

3. Comparing the predictions for the two situations, one would expect a smaller latency decrease from dissolidarity to solidarity for high-anxious and high-neurotic *S*s as compared with their respective counterparts (interaction effect).

Predictions for the frequency of shift of opinion (R) in the dissolidarity situation were

[6] This derivation deviates from the general formula on page 163 which defines the strength of response tendency relevant to the task. The reason lies in the definition of the measure as latency of *any* response *S* may make first. Therefore defensive (evasive) statements represent an additional response tendency available to *S* and its habit strength was added to that of *S*'s opinion (**7**, p. 60). The habit strength of the opposite (nondominant) opinion, which should be very small at the beginning and not very different for high- and low-neurotic *S*s, was disregarded, but should be subtracted if it were of any importance. On the other hand, difference in latency of statement of *opinion*, as well as other measures, were predicted according to the general formula.

TABLE 2

MEAN RESPONSE LATENCY (IN SECONDS) AND MEAN PARTICIPATION QUOTIENT (PQ)

	Solidarity			Dissolidarity		
	Anxiety			Anxiety		
	Low	High	Difference	Low	High	Difference
Latency						
Neuroticism						
High	8.2	11.8	−3.6	117.0	54.6	62.4**
Low	10.9	8.5	2.4	122.6	94.5	28.1
Difference	2.7	−3.3		5.6	39.9*	
PQ						
Neuroticism						
High	.34	.42	−.08*	.207	.252	−.045**
Low	.42	.42	.00	.240	.220	.020
Difference	.08*	.00		.033	−.032	

* $P < .05$ (1 tail).
** $P = .005$ (1 tail).

derived as follows. When two responses are possible, that response occurs whose reaction potential (E) is stronger (**7**). Two cases should be considered: (*a*) If the reaction potential (E) of DRT of high-anxious (high D) *Ss* is represented by a curve lying higher than that of low-anxious (low D) *Ss* (**7**, p. 34), after extinction treatment, it should remain stronger in spite of a larger drop due to higher drive. Therefore the rate at which NRT becomes stronger than DRT should be higher for low-anxious than high-anxious *Ss* (**8**, p. 31). (*b*) After some extinction of $\bar{H}$ and reinforcement of $\bar{H}$, the reaction potential (E) of DRT of high-neurotic *Ss* should decrease less than that of low-neurotic *Ss*, because an E associated with high $\bar{H}$ should decrease less than E associated with low $\bar{H}$ (**7**, p. 30), and also because of stronger interference of defenses with NRT in high-neurotic *Ss*. Thus the rate at which NRT becomes stronger than DRT should be higher for low-neurotic than for high-neurotic *Ss*. Therefore within a given length of sessions, fewer shifts of opinion are predicted for both high-anxious and high-neurotic *Ss* than for their respective counterparts. This is in line with the earlier prediction of diminished responsiveness of these *Ss* to change in the situation.

RESULTS

Response latency.—Latency correlations between *Ss* of different categories, viz. high-low anxious and high-low neurotic, within each situation were not significant, but the r (.35) based on the performance of the same *Ss* in the two situations was significant. Since difference in variability between the two situations could not be removed by any transformation of variable known to me, each situation was analyzed separately, applying the square-root transformation and the standard analysis of variance within it (Tables 2 and 3), and testing differences between situations by a non-parametric analysis of variance (Table 4). Significance of some differences within rows and columns was also calculated by the t test and is given in Table 2.

TABLE 3

ANALYSES OF VARIANCE FOR RESPONSE LATENCY WITHIN SOLIDARITY AND DISSOLIDARITY SITUATIONS

Source	*df*	*MS*	*F*	*P*
Within Solidarity				
Anxiety	1	.15		
Neuroticism	1	.04	1/45.5	>.05
A × N	1	2.94		
Error	60	1.82		
Within Dissolidarity				
Anxiety	1	77.38	5.01	<.05
Neuroticism	1	20.97		
A × N	1	22.12		
Error	60	15.44		

TABLE 4

NONPARAMETRIC ANALYSES OF VARIANCE OF RESPONSE LATENCY DIFFERENCES BETWEEN SOLIDARITY AND DISSOLIDARITY *

Source	*df*	*H*	*P*
Solidarity	1	71.04	<.01
S × A	1	11.64	<.01
S × N	1	6.96	<.01

* Kruskal and Wallis Test (15, p. 436).

Table 4 gives the results of three independent nonparametric analyses of variance (**15**): data for the main effect of solidarity were the raw latencies in the two situations. Data for S × N and S × A were the differences between the two situations for each *S*, so that the significance figures stand for the significance of second-order differences, i.e., first-order interactions. Higher order interactions were, of course, lost.

As expected, in the solidarity situation *Ss* showed better and more even performance than in the dissolidarity situation. In the latter, although the expected main effect of anxiety was significant, the difference is due mainly to a much shorter latency of those high-anxious *Ss* who also had a high neurotic score. These *Ss* also improved in the solidarity situation less than the others. Their response was, as predicted, stronger but somewhat less adaptable toward a changing stimulus situation than that of low-anxious, high-neurotic *Ss*. However, since the interaction effect was not significant, further evidence is needed before this conclusion is accepted as confirmation of our hypothesis. *Mutatis mutandis*, the same applies to high-neurotic, high-anxious *Ss*. But the main effect of neuroticism fell short of significant level.

Degree of participation (PQ).—Mean PQ in the solidarity situation (.40) was significantly larger than in the dissolidarity situation (.25), with $F = 18.2$ and $P < .01$. The predicted difference between mean PQ for high- and low-anxious *Ss* in the dissolidarity situation was almost significant for the high-neurotic category: $P \sim .04$ (1 tail). No other differences were significant, but some were in the predicted direction.

The degree of participation was also analyzed in another way: individual participation data were plotted cumulatively against duration of the discussions, in steps of 1 min. The curves obtained in both situations were linear and therefore PQ not only measures the total participation, but also its rate throughout the discussion.

Frequency of shift of opinion in dissolidarity situation.—Frequency data were analyzed by a formula originated by Dr. R. W. B. Jackson for a 3×3 matched-pair entry table. The formula for matched pairs was used because we compared pairs of *Ss* who had the same assistant pairs in their groups. Since correlations were not significant, however, the results were also checked by a formula for uncorrelated entry table, but remained unchanged. Significant differences were also found between high- and low-anxious *Ss* who had a high-neurotic

TABLE 5

FREQUENCY OF SHIFT OF OPINIONS AMONG *Ss* WITH LOW AND HIGH NEUROTICISM SCORES (N = 32 pairs of *Ss*)

	High Neuroticism			Totals
	0	1	2	
Low Neuroticism				
0 = no shift	6	2	0	8
1 = qualified shift	5	2	2	9
2 = complete shift	7	3	5	15
Totals	18	7	7	32
Probability ($z = -2.84$)				.003

score: $z = -2.05$, $P = .02$ (1 tail). These results confirm the expectation that fewer high-neurotic than low-neurotic *Ss* would shift their opinions, but only partly confirm the other hypothesis, that fewer high-anxious than low-anxious *Ss* would shift their opinions, since this is true only of those high-anxious *Ss* who also had a high-neurotic score.

Discussion

Validity of results obtained, of course, depends on the validity of the assumptions made, especially those concerning the values of variables involved, and on the successful control of the habit strength of opinions. Failure to obtain confirmation of some predictions concerning the degree of participation may be explained by a possible operation of Type II error, and the interaction effect of the two personality variables.

It has been objected that the difference between mean PQ in the two situations may be due to assistants' rather than to *Ss*' behavior. There is only partial truth in this criticism: *S* was not forced to speak in the solidarity, nor was he barred from speaking in the dissolidarity situation; dissolidarity sessions, however, were longer than solidarity sessions, and this difference may be due to assistants' behavior. The *Ss*' mean PQ, adjusted for this difference, is .30; the difference between the mean PQ's in the two situations remains significant ($t = 4.8$, $P < .01$).

Since predictions based on the definition of anxiety as a more or less chronically high emotional responsiveness were confirmed, this definition should be retained. In order to account for some recent findings (**2, 6, 12**), which seem to be adverse to this interpretation of anxiety, one should take into consideration experimental conditions, which may increase or reduce drive, and the magnitude of differences in anxiety, for one cannot expect large differences in performance when differences in anxiety levels are insignificant. These considerations perhaps can remove the basis of disagreement between Taylor, Spence, and Farber (**12**) one one hand, and Hilgard, Jones, and Kaplan (**6**) on the other.

Concerning the two relatively invariant forms of behavior obtained in this study, the following tentative conclusions can be stated: (*a*) lack of response adaptability to a changing stimulus may be due to a high level of motivation and neuroticism combined; (*b*) persistence of an unchanging response in the face of an unchanging aversive stimulus may be due to high neuroticism alone.

The present and previously obtained evidence (**3**) is encouraging for further experimental exploration of the generalized S-R theory as a tool for conceptualization of the structure of social situations, and for derivation of hypotheses. Experimentation is going on in our laboratories.

Finally, a word about *Ss*' self report on the effects of the experiment. A questionnaire was sent to each *S* asking how he or she had been affected by the experiment. The results may be summed up as follows: (*a*) *Ss* with high anxiety or neuroticism scores (who showed the least amount of change in their behavior during the experiment) mostly reported that they had felt "upset" by it; (*b*) *Ss* with low anxiety or neuroticism scores (whose performance showed more effect of the experiment) mostly reported that they had enjoyed it. The differences were significant for high-low anxiety *Ss* ($\chi^2 = 6.2$, $P < .05$) and were in the right direction ($\chi^2 = 1.79$, $P > .05$) for high low neuroticism *Ss*. This finding may have some bearing on psychotherapy.

Summary

The problem of the relation to behavior of three variables—anxiety, neuroticism, and group solidarity—was investigated. The experiment was designed and predictions derived with the aid of a generalized behavior theory based on Hull and Pavlov. In the resulting 2^3 factorial experiment, the two personality variables were varied by selection of *Ss* with extreme anxiety and neuroticism scores, the social variable was varied with the aid of role-playing assistants. All *Ss* received the solidarity (social approval) and dis-

solidarity (social disapproval) treatments in groups of three, made up of one *S* and two assistants. Among the findings may be mentioned the bettering and leveling effect of solidarity on performance, as against the lowering and differentiating effect of dissolidarity; greater rigidity of opinion of high-neurotic *S*s in the face of an unchanging aversive stimulus; shorter latency of response of high- than low-anxious and neurotic *S*s in a dissolidarity situation, their lower adaptability to a changing stimulus; and, finally, indication of usefulness of the generalized behavior theory as a tool for construction of experiments and derivation of hypotheses concerning social situations.

References

1. Beall, G. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, 1941–2, **32**, 243–262.
2. Bindra, D., Paterson, A. L., & Strzelecki, J. On the relation between anxiety and conditioning. *Canad. J. Psychol.*, 1955, **9**, 1–6.
3. Cervin, V. Experimental investigation of behavior in social situations: I. Behavior under opposition. *Canad. J. Psychol.*, 1955, **9**, 107–116; II. Individual behavioral effects of change in group attitude from opposition to cooperation. *Canad. J. Psychol.*, 1955, **9**, 155–160.
4. De Valois, R. L. The relation of different levels and kinds of motivation to variability of behavior. *J. exp. Psychol.*, 1953, **47**, 392–398.
5. Eysenck, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.
6. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
7. Hull, C. L. *Essentials of behavior*. New Haven: Yale Univ. Press, 1951.
8. Hull, C. L. *A behavior system*. New Haven: Yale Univ. Press, 1952.
9. Maier, N. R. F. The premature crystallization of learning theory. In *Learning theory, personality theory, and clinical research: the Kentucky symposium*. New York: Wiley, 1954.
10. Silverman, R. E. Anxiety and the mode of response. *J. abnorm. soc. Psychol.*, 1954, **49**, 538–542.
11. Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
12. Spence, K. W., Farber, I. E., & Taylor, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.
13. Taylor, J. A., & Spence, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–502.
14. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
15. Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

(Received for early publication
August 12, 1955)